Running Head: High throughput phenotyping of maize

Corresponding Author: Jianbing Yan and Wanneng Yang

National Key Laboratory of Crop Genetic Improvement and National Center of Plant Gene Research (Wuhan), Huazhong Agricultural University, Wuhan, 430070, China
Tel: +86-27-87280110, Fax +86-27-87280016
E-mail: Jianbing Yan (yjianbing@mail.hzau.edu.cn) and Wanneng Yang (ywn@mail.hzau.edu.cn)

Research Area: Breakthrough Technologies
High-throughput phenotyping and QTL mapping reveals the genetic architecture of maize plant growth

Authors:
Xuehai Zhang1,5, Chenglong Huang2,5, Di Wu2, Feng Qiao1, Wenqiang Li1, Lingfeng Duan2, Ke Wang2, Yingjie Xiao1, Guoxing Chen3, Qian Liu4, Lizhong Xiong1, Wanneng Yang1,2,* & Jianbing Yan1,*

1National Key Laboratory of Crop Genetic Improvement, National Center of Plant Gene Research (Wuhan), Huazhong Agricultural University, Wuhan 430070, PR China
2College of Engineering, Huazhong Agricultural University, Wuhan 430070, PR China
3MOA Key Laboratory of Crop Ecophysiology and Farming System in the Middle Reaches of the Yangtze River, Huazhong Agricultural University, Wuhan 430070, PR China
4Britton Chance Center for Biomedical Photonics, Wuhan National Laboratory for Optoelectronics, and Key Laboratory of Ministry of Education for Biomedical Photonics, Department of Biomedical Engineering, Huazhong University of Science and Technology, Wuhan 430074, PR China
5These authors contributed equally to this work.

Summary:
Combining high-throughput phenotyping facility and large scale QTL mapping dissects the dynamic genetic architecture of maize development by using a RIL population.
Footnotes:

Author contributions
W.Y. and J.Y. supervised the project. X.Z., C.H., W.Y. and J.Y. designed the research, performed the experiments, analyzed the data and wrote the manuscript. D.W., F.Q., W.L., L.D., K.W., Y.X. and G.C. performed the experiments. Q. L. and L.X. contributed for HRPF development.

Acknowledgements
This research was supported by the National Key Research and Development Program of China (2016YFD0100103), the National Program on High Technology Development (2013AA102403), the National Natural Science Foundation of China (31600287, 31525017), Hubei Provincial Natural Science Foundation of China (No. 2015CFB529), and the Fundamental Research Funds for the Central Universities.

Competing financial interests
The authors declare no competing financial interests.

* Corresponding author; e-mail Jianbing Yan (yjianbing@mail.hzau.edu.cn) and Wanneng Yang (ywn@mail.hzau.edu.cn).

5 These authors contributed equally to this work.
Abstract

With increasing demand for novel traits in crop breeding, the plant research community faces the challenge of quantitatively analyzing the structure and function of large numbers of plants. A clear goal of high-throughput phenotyping is to bridge the gap between genomics and phenomics. In this study, we quantified 106 traits from a maize recombinant inbred line population (n=167) across 16 developmental stages using the automatic phenotyping platform. QTL mapping with a high-density genetic linkage map, including 2496 recombinant bins, was used to uncover the genetic basis of these complex agronomic traits and 988 QTLs have been identified for all investigated traits, including 3 QTL hotspots. Biomass accumulation and final yield were predicted using a combination of dissected traits in the early growth stage. These results reveal the dynamic genetic architecture of maize plant growth and enhance ideotype-based maize breeding and prediction.
Maize (Zea mays L.) is one of the most widely grown crops worldwide and is not only a staple food for people and animals, but also an important industrial material for fuel and many other uses. Maize has served as a model plant with distinct advantages, including its levels of phenotypic and genetic variation (Yan et al., 2011). In the past century, maize yield has increased eightfold due to the efforts of plant breeders who harnessed genetic variations to breed for improvements in desired traits (Haley, 2011). It is predicted that 9 billion people living on this planet by 2050 will require 70% more food than today’s population and that more than half of the increased demand for cereals will come from maize (Yan et al., 2011). A large gap exists between current yield increase in global cereal production and the predicted demands for the next few decades. Substantial changes in breeding technologies for agronomic processes and crop improvement will be required (Tester and Langridge, 2010).

With the rapid development of next generation sequencing and high-density SNP genotyping technologies, linkage mapping and genome-wide association studies (GWAS) have been widely used to dissect the genetic architecture of agriculturally important traits in commercial maize (Buckler et al., 2009; Kump et al., 2011; Tian et al., 2011; Li et al., 2013; Wen et al., 2014; Wen et al., 2015). Many genes and variants underlying agriculturally important traits have been discovered in crops (Kesavan et al., 2013; Zuo and Li, 2013; Martinez et al., 2016). However, precision phenotyping still remains a bottleneck (Furbank and Tester, 2011). Traditional phenotyping is usually labor-intensive, time consuming, lower throughput, costly, and frequently destructive to plants (Chen et al., 2014), and is far behind the development of other -omic studies such as genomics, though efforts have been made to improve phenotyping efficiency (Yang et al., 2013).

'Plant phenomics' has been defined as the nondestructive and accurate acquisition of high-dimensional phenotypic data on an organism-wide scale across plant development (Houle et al., 2010). Recently, some high-throughput plant phenotyping platforms (Reuzeau et al., 2005; Nagel et al., 2012; Honsdorf et al., 2014) and...
open-source image analysis pipelines (Hartmann et al., 2011; Chen et al., 2014; Klukas et al., 2014) were developed to quantify phenotypic traits at the population level for different plant species. High-throughput non-invasive phenotyping has also been successfully adopted to assess the genetics of estimated biomass dynamics in maize (Junker et al., 2014; Muraya et al., 2016). In a previous work, an automatic phenotyping platform (RAP) was designed to achieve high-throughput screening of rice plants for genetic studies (Yang et al., 2014). In the present study, the RAP has been expanded for high-throughput phenotyping of maize plants. The dynamic growth phenotype of a recombinant inbred line (RIL) population containing 167 recombinant individuals, was measured from seedling to tasseling stage with 16 time points. 106 different traits were obtained. That, combined with the ultra-high density linkage map, including 2496 recombinant bins, allowed us to perform large scale QTL mapping. In total, 988 QTLs, including three hotspots, were identified, which has provided useful information for future maize genetic improvement and helps us understand the dynamic genetic architecture of plant development and growth.

Results

High-throughput plant phenotyping platform for maize

We cultivated the maize RIL population under greenhouse conditions (Figure 1a) and using the RAP platform phenotyped each individual plant (Figure 1b) from the seedling to tasseling stage at 16 time points (T1-T16, also see the Supplementary Video 1 for the process). At each time point, 15 side-view images and 1 top-view image of each individual plant were taken at the inspection unit using a color imaging camera (Stingray F-504C, Allied Vision Technologies, Germany). Two replicates of the RIL population and the two parents were inspected once every three days from seedling to tasseling stage, yielding ~476 GB of data. With the modified image analysis pipeline (Figure 1c), 106 phenotypic traits including 10 plant morphological traits, 22 leaf architecture traits, 1 plant color trait, 3 biomass related traits, 6 histogram texture traits, and 64 growth related traits, were extracted (Figure 1d). The definitions of these 106 traits are shown in Supplementary Table 1 and...
**Supplementary Note 1.** The time costs of plant screening and image analysis for each plant were 45s and 10s, respectively. The operating procedure for the RAP-maize is provided in Supplementary Video 1, and details of the image analysis pipeline are shown in the Methods and in Supplementary Figure 1.

**Performance evaluation of the RAP-Maize**

To model the biomass and evaluate the measuring accuracy of the RAP-maize, a subset of the maize association mapping panel (Yang et al., 2011; Li et al., 2013) (387 individuals from 100 genotypes with 3-4 replicates each, see Methods) was grown separately and measured for 3 traits automatically and manually in 4 stages. In the previous study (Yang et al., 2014), the side projected area (SA) showed good correlation with manually measured fresh weight (FW) and dry weight (DW) with R^2 > 0.79. To determine the best model for predicting the FW and DW, 10 models (including linear, quadratic, exponential, and power models) were evaluated using the adjusted coefficient of determination (adjusted R^2) and the mean absolute percentage error (MAPE) as the index, the statistical details are shown in Supplementary Table 2. The results showed that the power model (Mode 9 and Mode 10) had a higher R^2 and less MAPE for the FW. Finally, Mode 9 with one variable SA_{max} was selected as the best model for FW modeling, which was one of the reasons why the image with maximum SA was automatically selected for the following image analysis and traits extraction. A similar result was observed with the same model for DW (Supplementary Table 3). Scatter plots showed that the R^2 were greater than 0.97 between manual versus automatic measurements for plant height (Figure 2a), FW (Figure 2b), and DW (Figure 2c). These results demonstrate that automatic measurements are as good as manual measurements but with higher throughput.

**Natural variation of phenotypic traits and heritability**

The RIL population manifested high diversity for most of the 106 investigated phenotypic traits at each time point (Supplementary Data 1), ranging from 1.07 to 5.56 and 1.02 to 14.95-fold change at minimum and maximum level at 16 time points,
respectively. For all investigated phenotypic traits, an approximately 3-fold change was observed on average at different time points (Table 1).

The investigated traits showed greater than 0.5 heritability for most traits at most of the time points (Figure 3, left) as exemplified by the natural plant height (NPH) and one of morphology related trait digital volume (Figure 3, top right). For most traits, $H^2$ estimates were low in the six early developmental stages (i.e., 0.064 for U_TEX (the uniformity, one of six histogram texture traits) at first time point and 0.15 for LTA at sixth time point) and increased with the growth and development of maize in the late developmental stages. Some traits, including LTA, SDLTA, FDIC, FDNIC, and LSA_above, have low $H^2$ across all developmental stages (Figure 3, left), which might be due to the lower genetic variation of these traits. Some traits, including PP, PBR, MPH, NPH, and LSA, showed higher heritability in late developmental stages, which may due to the genes controlling these traits expressed primarily in the late stages or environmental effects are larger at early stages, but got averaged over time in later stages (Figure 3, bottom right).

**QTL analysis**

In this study, 167 of the genotyped RILs were used for QTL mapping for all the investigated phenotypic traits at each time point. In total, 42 to 82 QTLs were identified at each time point (Table 1). The number of QTLs for each trait ranged from 1 to 8, with a mean of 1.7 to 2.7 across the 16 time points. In total, 938 QTLs were identified for 42 investigated phenotypic traits across 16 time points of maize growth. The percentage of phenotypic variation that each QTL could explain ranged from 5.5% to 26.6%, with a mean ranging from 8.7% to 10.5% at the 16 time points, respectively (Table 1). For the growth rate related traits DW and FW, across 16 time points, a total of 50 QTLs were detected. The number of QTLs for each growth rate related trait ranged from 1 to 4, with a mean of 1.5 to 1.9. The percentage of phenotypic variation that each QTL could explain ranged from 7.3% to 17.8%, with a mean ranging from 8.4% to 10.0% for these 64 growth rate related traits (Supplementary Table 4). For these 988 QTLs, the mapping resolution (QTL support
interval) ranged from 0.3 cM to 10.8 cM, with a mean of 2.7 cM (~3.9 Mb) (Supplementary Figure 2). Furthermore, these QTLs can be categorized into 152 non-redundant QTLs (Supplementary Data 2), which may be due to the highly positive or negative correlations between paired traits (Supplementary Figure 3).

The detailed information, including location, peak marker, additive effect, QTL support interval and explained phenotypic variance ($R^2$) of each QTL for each trait, is shown in Supplementary Data 2. A traits-locus network, including all traits across 16 development stages and their corresponding significant loci, shows the complex relationships among traits and detected loci. Some obvious connection nodes were found which could correspond to factors regulating maize growth (Figure 4a). QTL distribution across chromosomes (Supplementary Figure 4) was not random, three QTL hot spots were observed across the maize genome, on chromosome 3, 7, and 10 (Figure 4b). Moreover, the following observations were made: (1) A single QTL affecting a particular trait mapped at several time points of development, for example, a QTL affecting DW located at 120.5 cM on chromosome 7 was detected at 7 time points (Figure 4c) and a QTL for TLL on chromosome 2 was detected at 10 time points (Supplementary Figure 4). This implies that the gene affecting these traits may have expressed in early stages. (2) There was some overlap between QTLs affecting DW and SA (Figure 4d and 4e) since high correlations were observed between the two traits (Supplementary Table 3), indicating that SA can replace DW in QTL analysis in maize. (3) At a particular stage, most traits were controlled by a number of QTLs with minor to moderate effects (Supplementary Figure 4 and Supplementary Data 2).

Recently, large-scale metabolic trait QTL mapping was also performed in the same RIL population in different tissues including leaf (Wen et al., 2015). A common QTL hotspot was observed on chromosome 10 (Supplementary Figure 5) in the metabolic traits QTL mapping and in the present study, indicating that this genome region may not only affect variation of most metabolic traits but also control maize growth traits.
Prediction of digital biomass accumulation

It would be very helpful for maize breeding if we could use the phenotypic data obtained in the early growth stage to predicate the final biomass or yield. The digital biomass (side projected area, SA) of the 387 individuals was calculated and showed good correlation with the manually measured FW and DW (Supplementary Table 2 and Supplementary Table 3), and thus can be used to represent the biomass. After the digital biomass of the RIL population at 16 different time points was obtained (Figure 5a), we tested the 6 models (linear, power, exponential, logarithm, quadratic, and logistic models) for their ability to predict digital biomass in the early growth stage. The digital biomass measurements at 16 time points were divided into a training set and a testing set. For example, if the number of time points for the training set was 6 (T1-T6), the corresponding number of time point for the testing set was 10 (T7-T16). The results were evaluated by comparison of $R^2$, MAPE, and SD$_{APE}$ values. As shown in Supplementary Figure 6, when the number of time points for the training set is large enough (such as 11), the prediction results of the power, exponential, quadratic, and logistic models were all satisfactory ($R^2$ was 0.96, the MAPE and SD$_{APE}$ were both below 30%). However, when the number of time points for the training set was 6, only the exponential model with the testing set showed good prediction ability ($R^2$ was 0.96, the MAPE and SD$_{APE}$ were both below 20%). The comparison of actual digital biomass and predicted digital biomass is shown in Figure 5b), and indicates that from the seedling stage to the tasseling stage, the exponential model had better prediction ability for digital biomass accumulation, even in the early growth stage (Supplementary Figure 6 and Supplementary Table 5).

A number of novel traits could be used as indicators for final yield prediction

It would be important for plant breeding if we could use the measured traits, especially the traits measured in early development stages, to predict the final grain yield. To test whether this is possible, the variance explained of grain yield with different traits in different stages was evaluated (Supplementary Table 6). Up to 54.6% of the phenotypic variance of grain yield could be explained by combining 16 traits.
across all 16 time points. If only 8 phenotypic traits at 4 time points (T1, T8, T9, T16) were used, 29.6% of the grain yield variance was explained (Figure 5c). Based on the values of coefficients in the selected model for grain yield (Supplementary Table 7), we established an ideotype maize plant (Figure 5d). The grain yield had (1) positive correlation with the leaf morphological traits (FDIC_1) \((r = 0.261)\) and the leaf angle in the upper half of the plant (LTA_above_1) \((r = 0.133)\). This implies that a more wavy-shaped leaf may maximize the area receiving light in the seedling stage; (2) a positive correlation with the leaf angle in lower half of the plant (LTA_below_9) \((r = 0.156)\) and a negative correlation with leaf angle in upper half of the plant (LTA_above_9) \((r = -0.152)\). This distribution of leaf architecture can maximize the light harvesting area, thus increasing photosynthetic efficiency; (3) a positive correlation with the SDLC_8 (standard deviation of leaf curvature per plant) \((r = 0.171)\); higher SDLC indicates higher variability of leaf angles within a plant; (4) a positive correlation with the GCV_8 \((r = 0.114)\); higher GCV means more dark green leaves, which may be related to greater chlorophyll content; (5) a positive correlation with leaf length in the upper half of the plant (LNL_above_16) \((r = 0.192)\). These results were consistent with the smart canopy concept for maize plant, which promises to maximize the potential for light harvesting per unit of land area (Ort et al., 2015). Interestingly, a QTL hot spot was identified on chromosome 10 for SDLC, which overlapped with a metabolic trait QTL (Wen et al., 2015) (Supplementary Figure 4 and Supplementary Figure 5). Cloning of this QTL may be helpful for understanding plant architecture regulation and the associations with grain yield.

Discussion

Due to the limitations of traditional phenotyping, which is labor-intensive, time-consuming, low throughput, and costly (Chen et al., 2014), most previous QTL studies focused on a limited number of traits, usually at the mature stage (Yan et al., 2003; Osman et al., 2013; Zhang et al., 2013). Plant growth is a dynamic process and the timing of end-point measurement will greatly influence the outcome of mapping (El-Lithy et al., 2004). In this study, using a modified image analysis pipeline, the rice
automatic phenotyping platform (RAP) was expanded for use in a maize RIL population for high-throughput quantifying of multiple traditional and novel features at 16 development stages accurately ($R^2 > 0.97$) and efficiently (45 seconds were needed for screening and image analysis per plant). Both traditional traits (i.e. plant height) and many novel features (i.e. FDIC, GCV, LTA) could be investigated non-destructively at multiple time points. RAP-Maize provides a good opportunity to study dynamic development process in maize and to better understand the underlying genetic mechanisms.

By combining the high-throughput maize phenotyping platform and the high-density linkage map, 988 QTLs were identified that provided the opportunity to investigate QTL distribution at the genome-wide level (Supplementary Data 2 and Supplementary Figure 4). The following observations were made: (1) QTLs were different at different stages, indicating that plant growth regulation mechanism changes over time; (2) Most traits were controlled by a large number of QTLs, consistent with the quantitative nature of these traits as well as with previously studied agronomic traits (Buckler et al., 2009; Xiao et al., 2016); (3) QTL distribution across chromosomes was not random (Figure 4b, Supplementary Figure 4) and three QTL hot spots were identified (Figure 4b) including one hotspot located in the genomic region previously identified for metabolic trait QTL (Wen et al., 2015), which could be the key node for regulating plant growth. In total, 53 candidate genes within the first three peak bins (91.96 Mb - 95.21 Mb) of this hotspot were identified (Supplementary Table 8) and great efforts were still required to find out the casual gene(s). Another QTL hot spot located on chromosome 7 associated with DW was identified across 7 time points (Figure 4c-e). In total, 28 candidate genes located in the first three peak bins (161.62 Mb - 162.16 Mb) (Supplementary Table 9), and only four candidate genes (GRMZM2G180490, GRMZM2G010702, GRMZM2G151649, and GRMZM2G057023) were located within the peak bin (161.95-162.04 Mb) of the QTL (Figure 4f). GRMZM2G180490 encodes an adenylyl-sulfate kinase; GRMZM2G010702 has an unknown function; GRMZM2G151649, the homologous of AT3G01400.1, encodes an ARM repeat superfamily protein, which is involved in
the ubiquitination pathway regulating the development of seed size in soybean (Xie et al., 2014); GRMZM2G057023 encodes an interferon-related developmental regulator and expresses highly in leaf (Supplementary Figure 7), which might be the most likely candidate gene affecting DW. Furthermore, for the biomass, significant loci detected here were not overlapped with other studies (Barrière et al., 2010; Riedelsheimer et al., 2012; Rincent et al., 2014; Muraya et al., 2016) and only one QTL located in the hotspot on chromosome 7 which was coincided with the fresh shoot biomass (biomass) and metabolite QTL identified previously (Wen et al., 2015). Cloning this QTL hotspot should help to explain an underlying mechanism of plant growth and metabolic regulation. In a recent study, automated non-invasive phenotyping method was also used to monitor the plant sizes of 252 diverse maize inbred lines by focusing on biomass at 11 different developmental time points and 12 main effect marker-trait associations were identified (Muraya et al., 2016). Association mapping in maize is a very powerful tool to identify the genes with high resolution if millions of molecular markers and large population size were used (Yan et al., 2011; Liu et al., 2016). Only few marker-trait associations were identified for biomass in the recent study (Muraya et al., 2016) since the marker density is low and sample size is median. We also compared the merits and demerits of high-throughput phenotyping technology in different plant species especially in crops (Supplementary Table 10). Our platform has obvious advantages for detected trait/QTL number, more importantly, we have detected many novel traits undetected previously which can be used for yield and biomass prediction. However, it was indoor shoot based phenotyping and need to expand to the field or filed plot level.

Predicting the crop yield by using simple phenotypic indicators available early in development would greatly aid maize breeding. Molecular markers have been widely used in breeding diagnostics and are especially efficient for traits controlled by major genes (Collard and Mackill, 2008). In the present study, we found that a few indicators in the early growth stage of maize could be used to predicate final grain yield. About 30% of the variance could be explained by using only 8 phenotypic traits at 4 time points. This is an impressive result, given that grain yield data came from 7
different field environments in different years and only a few measured traits from seedling to tassling stage were being used. These findings provide us useful clues for ideotype-based maize breeding by optimization of leaf and plant architecture, such as: (1) smaller to bigger leaf angles from top to ground; (2) more wave-like and dark green leaf; (3) longer leaf, especially around and up ears. We simplified the model as shown in Figure 5d. More importantly, most of the mentioned traits are very difficult to measure by the traditional method and now can be manipulated automatically in the early stage of maize growth. The smart canopy for maize is an integrated concept that still needs to be tested and proven in the field in the future. Combining the current field phenotyping platforms, such as the aerial sensing platform (Berni et al., 2009), ground-base field phenotyping at the plot level (Andradesanchez et al., 2013), and the movable imaging chamber in the field (Busemeyer et al., 2013), the robust image analysis pipeline could also be transferred to the field (Berni et al., 2009; Andradesanchez et al., 2013; Busemeyer et al., 2013) for high-throughput phenotyping. In summary, combining the high-throughput phenotyping technology and large-scale QTL analysis not only greatly expanded our knowledge of the maize dynamic development process, but also provided a new strategy for breeders to optimize plant architecture towards ideotype breeding in maize.

Methods

Plant materials, growth conditions, and experiment design

In the present study, a maize RIL population (Chander et al., 2008a; Chander et al., 2008b; Wen et al., 2015) with its parents (B73 and BY804) were planted in RAP (Yang et al., 2014) with two replications. All the maize RILs were screened at sixteen different developmental time points (once every three days starting from 22 to 67 days after sowing, DAS, e.g. T1 to T16 represent 22, 25, 28, 31, 34, 37, 40, 43, 46, 49, 52, 55, 58, 61, 64, 67 DAS, respectively), the sowing date and inspection dates were provided in Supplementary Table 11. The growth conditions were as follows: fertilizing was carried out at sowing, V5 and V9 stage, respectively (60-kilogram...
water + 370.68 gram carbamide + 330.76 gram potassium dihydrogen phosphate + 94.24 gram potassium chloride, to be fully dissolved, with 150 milliliter liquid fertilizer for each plant per time). A subset of the association mapping population (Yang et al., 2011; Li et al., 2013), including 100 diversity inbred lines, was randomly selected and sowed on the same day in phenotypic platform with four replications, screened using RAP and manually measured at 36, 48, 60, 70 DAS (Supplementary Table 11). Destructively measured traits were obtained for biomass modeling and the correlation between manual measurement and automatic measurement was calculated. The growth movie of the parents and 4 selected recombinant lines is shown in Supplementary Video 2~7.

The RIL population was also planted in Henan, Hubei, Chongqing, Yunnan, and Hainan provinces, China during 2011 and 2012. In the seven environments, at least five good open-pollinated ears were harvested from each row for measuring ear weight and cob weight. Yield data (ear weight minus cob weight) for each line was recorded. Best linear unbiased prediction (BLUP) was obtained by fitting the mixed liner model in R package lme4 (Team, 2013) for estimation of breeding value of each line across all environments and the BLUP values were then combined to reduce the prediction bias caused by the unbalanced data. Finally, the BLUP data of grain yield across 7 environments and all investigated phenotypic traits obtained across 16 time points were put into the grain yield prediction model. The experiment design was shown in Supplementary Figure 8.

Image analysis and traits extraction

Maize image analysis was carried out as shown in Supplementary Figure 1. First, the image with the maximum plant width was automatically selected from 15 different angles; then, the E x G component was extracted and the OTSU method (Ohtsu, 1979) was applied to acquire the binary image (Supplementary Figure 1b). Second, region growing algorithm was used to obtain the whole plant binary image (Supplementary Figure 1c). Using the binary image, the plant morphological traits
and projected area were calculated. Moreover, the color trait and histogram texture traits were computed by matching the binary image and original color image. Then, the parallel thinning algorithm was performed to create the skeleton image (Supplementary Figure 1d), and the Hough transformation was applied to distinguish the leaf skeleton from the stem skeleton (Supplementary Figure 1e). With this information, the stem length and total leaf were identified. Finally, each leaf skeleton was identified and labeled (Supplementary Figure 1f), and the traits for each leaf were calculated, including leaf angle, leaf length, and leaf curvature. With biomass (fresh weights, FW, and dry weights, DW) obtained at different time points, growth related traits were calculated. Detailed information for traits definition is shown in Supplementary Table 1, and details of traits extraction are described in Supplementary Note 1. The software interface and the source code of image analysis pipeline were added in the Supplementary Figure 9 and 10. In addition, the Labview programs and dynamic link library (DLL) enrolled in maize image analysis pipeline were listed in Supplementary Table 12. All the source code and software applications, including Labview programs, DLL, and cpp documents, can be downloaded using the link: http://plantphenomics.hzau.edu.cn/checkiflogin_en.action (Username: UserPP, Password: 20170108pp).

Biomass modeling and digital biomass accumulation modeling

To determine the best model for measuring FW and DW, 10 models (including linear, quadratic, exponential, and power models) were evaluated using the adjusted coefficient of determination (adjusted $R^2$) and the mean absolute percentage error (MAPE). The statistical analysis (mainly the linear stepwise regression) for biomass modeling was implemented with the SPSS software (Statistical Product and Service Solutions, Version 13.0, SPSS Inc., USA). After the predicted FW and DW at 16 different time points were obtained, maize plant growth was modeled with 6 models: linear, power, exponential, logarithm, quadratic, and logistic. In order to test the prediction ability of the different models in the early growth stage, the FW values at
16 time points were divided into two parts: a training set and a testing set. The results of trait fitting were evaluated by comparison of $R^2$, MAPE, and SDAPE values. The statistical analyses of the 6 models (linear, power, exponential, logarithm, quadratic, and logistic model) for maize plant growth were implemented with LabVIEW 8.6 (National Instruments, Inc., USA) and MATLAB 2011 (the Mathworks, Inc., USA).

**Grain yield prediction using plant phenotypic traits in the early growth stage**

To evaluate the variance explained for the maize grain yield in the early growth stages, the linear stepwise regression was used with maize plant phenotypic traits. The entry value of use probability of F in the stepping method criteria was 0.05, and the removal value of use probability of F was 0.10. The variable was added into the model if F value was less than the entry value; however, the variable was removed if F value was higher than the removal value. The linear stepwise regression analysis for grain yield was implemented with the SPSS software (Statistical Product and Service Solutions, Version 13.0, SPSS Inc., USA).

**Heritability analysis**

Heritability ($H^2$) was calculated for each trait as follows:

$$H^2 = \frac{\sigma_G^2}{\sigma_G^2 + \sigma_e^2 / r}$$

Where $\sigma_G^2$ is the genotypic variance, $\sigma_e^2$ is the error variance, $r$ is the number of replications. The estimates of $\sigma_G^2$ and $\sigma_e^2$ were analyzed by analysis of variance (ANOVA) using the lmer function in lme4 package in R environment (Team, 2013) (version 3.1.3, R Foundation for Statistical Computing, http://www.r-project.org/).

**Genotype, linkage map construction, and QTL Mapping**

The genotypic data for the RIL population obtained from a former study (Wen et al., 2015), showed that a linkage map was 1790.2 centimorgans (cM) in length, including 2496 recombinant bins, 0.72 cM per bin on average. Details about the map construction and its description were reported in previous studies (Wen et al., 2015; Pan et al., 2016). In summary, the RIL population was genotyped using the Illumina
MaizeSNP50 BeadChip, which contains 56,110 SNPs (Ganal et al., 2011). SNPs with both missing rate and heterozygosity of less than 10% were used to construct the genetic linkage map which contains 2496 recombinant bins. In this study, 167 of the genotyped RILs were used for QTL mapping, using the seeding emergence and growth phenotypes. QTL analysis was performed by the composite interval mapping (CIM) method (Zeng, 1994), using the software Windows QTL Cartographer V2.5 (Wang S, 2007), for each investigated phenotypic trait and growth related traits at 16 time points. Walking step was set to 0.5 cM and zmap (model 6) with a 10 cM window was used. The bins or genetic blocks (a genomic region in which no recombination exists) were clearly defined and a uniform LOD value was assigned for each bin. For determining LOD threshold, 25 trait-time point combinations from all 736 trait-time point combinations were selected randomly for permutation tests with 500 times (P = 0.05). The results indicated that the LOD threshold ranged from 3.23-3.65, with a mean of 3.42. To simplify it, a threshold of LOD $\geq$ 3.5 was used to establish the presence of a QTL for all the traits. A QTL support interval was defined as the one-LOD drop position ranging from the QTL peak. All QTLs with overlapping QTL support interval were categorized as a non-redundant QTL. Possible candidate genes were identified within the QTL hot spot on chromosome 7 and 10 based on the filtered working gene list of maize genome downloaded from MaizeGDB (http://www.maizegdb.org). Candidate genes were annotated according to InterProScan (http://www.ebi.ac.uk/interpro/scan.html).

**Construction of the trait-locus network**

The nodes of the traits-locus network contain all traits across 16 development stages and their corresponding significant loci with the threshold value of LOD $\geq$ 3.5. All traits were labeled blue, all loci were labeled pink (Figure 4a). The network was visualized using the software Cytoscape version 2.6.3 (Shannon et al., 2003).
Supplemental Items

Supplementary Data 1 Variation of Investigated Phenotypic Traits in BY804/B73 Population and Two Parents at Different Stages.

Supplementary Data 2 QTL Information Summary of All Traits across 16 Time Points.

Supplementary Video 1 The operating procedure for the RAP-maize.

Supplementary Video 2 The growth movie of the parent B73.

Supplementary Video 3 The growth movie of the parent By804.

Supplementary Video 4 The growth movie of the recombinant line BB048.

Supplementary Video 5 The growth movie of the recombinant line BB054.

Supplementary Video 6 The growth movie of the recombinant line BB078.

Supplementary Video 7 The growth movie of the recombinant line BB096.

Supplementary Information: contains Supplementary Figures 1-10, Supplementary Tables 1-12 and Supplementary Notes 1

Supplementary Figures 1-10:

Supplementary Figure 1 Maize image analysis and traits extraction.

Supplementary Figure 2 Distribution of QTL mapping resolution.

Supplementary Figure 3 Correlation coefficients between paired traits for 42 traits investigated at 16 time points.

Supplementary Figure 4 Chromosomal distribution of identified QTLs with 42 primary phenotypic traits and 64 growth related traits.

Supplementary Figure 5 Comparison of heat maps for QTLs density between metabolic and investigated phenotypic traits in By804/B73 recombination population.

Supplementary Figure 6 Predication ability comparison of 6 models for digital biomass accumulation.

Supplementary Figure 7 The RNA-seq gene atlas for four genes (GRMZM2G180490, GRMZM2G010702, GRMZM2G151649 and GRMZM2G057023).

Supplementary Figure 8 Experimental design.

Supplementary Figure 9 The image analysis interface designed in the study.
Supplementary Figure 10: The flow chart of the program. The number represents the processing module of the following figure 1~10.

Supplementary Tables 1-12:

Supplementary Table 1: The 106 traits classification and abbreviation.

Supplementary Table 2: Statistical summary of the 10 developed models for fresh weight estimation (sample size = 387).

Supplementary Table 3: Statistical summary of the 10 developed models for dry weight estimation (sample size = 387).

Supplementary Table 4: Summary of QTL for Growth Rate Related Trait Identified at Sixteen Time Points.

Supplementary Table 5: Statistical summary of the 6 developed models for digital biomass accumulation (167 samples × 16 time points).

Supplementary Table 6: Detecting the phenotypic traits (not include growth related traits) significantly associated with yield (Tons per hectare) and calculating the percentage of the phenotypic variance explanation (R²).

Supplementary Table 7: The statistical details of coefficients in selected model for yield (time points: 1, 8, 9, 16 in the Supplementary Table 6).

Supplementary Table 8: Candidate genes and their annotations located in the first three peak bins in QTL hot spot located on chromosome 10.

Supplementary Table 9: Candidate genes and their annotations located in the first three peak bins in QTL hot spot located on chromosome 7.

Supplementary Table 10: Comparison of published work for combination of high-throughput phenotyping and QTL/GWAS analysis.

Supplementary Table 11: Experimental schedule of maize plant phenotyping.

Supplementary Table 12: The detailed information of Sub-vi and Dynamic link library (DLL) used in the study.

Supplementary Notes 1: Definition of the features.
Figure Legends

Figure 1 High-throughput maize plant phenotyping. (a) Maize plant sowing and maize cultivation, (b) The procedure of maize inspection, (c) The image analysis pipeline, (d) The extracted maize phenotypic traits and modeling.

Figure 2 Performance evaluation of RAP-maize. (a) Scatter plot of automatic measurements versus manual measurements for plant height, (b) Scatter plot of automatic measurements versus manual measurements for fresh weight, (c) Scatter plot of automatic measurements versus manual measurements for dry weight.

Figure 3 Phenotypic trait heritability during 16 growth stages. Heat map showing broad-sense heritability ($H^2$) of the investigated phenotypic traits over 16 time points (left), as exemplified by the digital volume for natural plant height (top right) and for mean $H^2$ (bottom right). Trait identifiers are given as in heat map and colored according to their classification as indicated. Black, plant morphological traits; Green, leaf architecture traits; Blue, biomass related traits; Pink, histogram texture feature; Purple, color trait. Mean, $H^2$ of mean value for all traits.

Figure 4 Chromosomal distribution of identified QTLs. (a) Network of QTLs associated with the traits obtained from RAP-maize. Blue nodes represent the 42 traits. Pink nodes represent all identified significant loci from QTL mapping. (b) Density of QTLs of all investigated phenotypic traits across the genome. The window size is 10 centimorgan (cM). Three QTL hotspots were marked with asterisks. Detailed information for all detected QTLs is shown in Supplemental Date 2. (c) Chromosomal distribution of dry weight (DW) QTLs identified across 16 time points, QTL regions (represented by the confidence interval for each QTL was assigned as 1-LOD drop of the peak) across the maize genome responsible for dry weight are shown as black solid rectangular boxes. QTL hot regions for dry weight on chromosome were marked with a gray hollow rectangular box. (d) LOD curves of QTL mapping for SA and DW around the chromosome 7 QTLs at 6th time point. (e) LOD values of the bins at the peak of QTL interval are shown as a function of their genetic positions. (f) Candidate genes (GRMZM2G180490, GRMZM2G010702, GRMZM2G151649, and GRMZM2G057023) located in the peak bin. Other candidates genes located in the two bins next to the peak bin are listed in Supplementary Table 9.

Figure 5. Predication of maize digital biomass accumulation and yield. (a) Heat map showing the difference of digital biomass accumulation with different maize recombinant inbred lines. (b) The comparison of actual digital biomass (blue line) and predicated digital biomass (red line, using digital biomass of the first 6 time points to predicate the digital biomass of the rest 10 time points); error bars represent the standard error of dry weight of 167 samples in each time point. (c) The scatter plot shows the relationship between the actual grain yield and predicated yield with the
predicted formula; a, b, c, d, e, f, g, h represent FDIC_1, LTA_above_1, GCV_8, SDLC_8, LTA_above_9, LTA_below_9, LNL_above_16, LSA_below_16, respectively; FDIC_1 and LNL_above_16 were leaf morphological traits; LTA_above_1, SDLC_8, LTA_above_9, LTA_below_9, and LSA_below_16 were leaf angle traits; GCV_8 was plant color traits; The black line was the fitting line and the standardized coefficients were shown in Supplementary Table 7; (d) The predicted ideotype maize plant based on the associated traits in early stages with higher GCV, FDIC, LNL_above, LTA_below, and lower LTA_above.

Table 1. Range, mean of fold changes, and summary of QTLs for investigated phenotypic traits identified at sixteen time points.

<table>
<thead>
<tr>
<th>Time point</th>
<th>No. of Traits (42)</th>
<th>Fold Change(RILs, Mean)</th>
<th>Fold Change(RILs, Range)</th>
<th>QTL number (Mean and Range)</th>
<th>No. of QTL (Mean and Range)</th>
<th>PVE (% Mean and Range)</th>
</tr>
</thead>
<tbody>
<tr>
<td>T1</td>
<td>29</td>
<td>3.06</td>
<td>1.10-10.03</td>
<td>50</td>
<td>1.7 (1-5)</td>
<td>10.5 (6.8-17.9)</td>
</tr>
<tr>
<td>T2</td>
<td>31</td>
<td>3.24</td>
<td>1.02-14.95</td>
<td>69</td>
<td>2.2 (1-5)</td>
<td>9.3 (6.9-14.7)</td>
</tr>
<tr>
<td>T3</td>
<td>31</td>
<td>3.02</td>
<td>1.01-12.06</td>
<td>59</td>
<td>1.9 (1-4)</td>
<td>9.4 (7.1-16.4)</td>
</tr>
<tr>
<td>T4</td>
<td>30</td>
<td>2.67</td>
<td>1.02-7.42</td>
<td>59</td>
<td>2.0 (1-5)</td>
<td>8.7 (7.0-16.0)</td>
</tr>
<tr>
<td>T5</td>
<td>29</td>
<td>2.95</td>
<td>1.03-9.31</td>
<td>66</td>
<td>2.3 (1-6)</td>
<td>9.4 (6.9-16.0)</td>
</tr>
<tr>
<td>T6</td>
<td>35</td>
<td>2.94</td>
<td>1.03-9.55</td>
<td>82</td>
<td>2.3 (1-5)</td>
<td>9.5 (6.6-19.4)</td>
</tr>
<tr>
<td>T7</td>
<td>36</td>
<td>2.97</td>
<td>1.04-10.83</td>
<td>69</td>
<td>1.9 (1-4)</td>
<td>9.8 (6.4-26.4)</td>
</tr>
<tr>
<td>T8</td>
<td>22</td>
<td>2.81</td>
<td>1.04-7.34</td>
<td>53</td>
<td>2.4 (1-5)</td>
<td>9.9 (6.2-26.6)</td>
</tr>
<tr>
<td>T9</td>
<td>30</td>
<td>3.11</td>
<td>1.08-10.61</td>
<td>65</td>
<td>2.2 (1-6)</td>
<td>9.6 (5.5-24.2)</td>
</tr>
<tr>
<td>T10</td>
<td>23</td>
<td>2.83</td>
<td>1.09-8.19</td>
<td>48</td>
<td>2.1 (1-8)</td>
<td>9.3 (6.8-26.3)</td>
</tr>
<tr>
<td>T11</td>
<td>27</td>
<td>2.70</td>
<td>1.09-8.00</td>
<td>50</td>
<td>1.9 (1-6)</td>
<td>9.7 (6.6-17.6)</td>
</tr>
<tr>
<td>T12</td>
<td>25</td>
<td>2.46</td>
<td>1.10-5.79</td>
<td>48</td>
<td>1.9 (1-5)</td>
<td>9.1 (6.4-15.3)</td>
</tr>
<tr>
<td>T13</td>
<td>22</td>
<td>2.58</td>
<td>1.08-7.69</td>
<td>42</td>
<td>1.9 (1-4)</td>
<td>10.0 (6.2-19.6)</td>
</tr>
<tr>
<td>T14</td>
<td>30</td>
<td>2.53</td>
<td>1.07-5.56</td>
<td>72</td>
<td>2.4 (1-6)</td>
<td>9.7 (6.1-17.7)</td>
</tr>
<tr>
<td>T15</td>
<td>27</td>
<td>2.55</td>
<td>1.06-6.28</td>
<td>56</td>
<td>2.7 (1-5)</td>
<td>10.1 (6.8-19.2)</td>
</tr>
<tr>
<td>T16</td>
<td>27</td>
<td>2.87</td>
<td>1.06-13.66</td>
<td>50</td>
<td>1.9 (1-5)</td>
<td>9.2 (6.7-15.2)</td>
</tr>
</tbody>
</table>

aNumber of traits with QTLs identified among the 42 measured traits in each time point in this study. b, c The values were calculated based on all 42 measured traits. dTotal number of QTLs identified at each time point. ePhenotypic variation explained (PVE) by each QTL.
References


Collard BC, Mackill DJ (2008) Marker-assisted selection: an approach for precision...


Kump KL, Bradbury PJ, Wisser RJ, Buckler ES, Belcher AR, Oropezarosas MA,


Zeng ZB (1994) Precision mapping of quantitative trait loci. Genetics 136:

Figure 1 High-throughput maize plant phenotyping. (a) Maize plant sowing and maize cultivation, (b) The procedure of maize inspection, (c) The image analysis pipeline, (d) The extracted maize phenotypic traits and modeling.
Figure 2 Performance evaluation of RAP-maize. (a) Scatter plot of automatic measurements versus manual measurements for plant height, (b) Scatter plot of automatic measurements versus manual measurements for fresh weight, (c) Scatter plot of automatic measurements versus manual measurements for dry weight.
**Figure 3 Phenotypic trait heritability during 16 growth stages.** Heat map showing broad-sense heritability ($H^2$) of the investigated phenotypic traits over 16 time points (left), as exemplified by the digital volume for natural plant height (top right) and for mean $H^2$ (bottom right). Trait identifiers are given as in heat map and colored according to their classification as indicated. Black, plant morphological traits; Green, leaf architecture traits; Blue, biomass related traits; Pink, histogram texture feature; Purple, color trait. Mean, $H^2$ of mean value for all traits.
Figure 4 Chromosomal distribution of identified QTLs. (a) Network of QTLs associated with the traits obtained from RAP-maize. Blue nodes represent the 42 traits. Pink nodes represent all identified significant loci from QTL mapping. (b) Density of QTLs of all investigated phenotypic traits across the genome. The window size is 10 centimorgan (cM). Three QTL hotspots were marked with asterisks. Detailed information for all detected QTLs is shown in Supplemental Data 2. (c) Chromosomal distribution of dry weight (DW) QTLs identified across 16 time points, QTL regions (represented by the confidence interval for each QTL was assigned as 1-LOD drop of the peak) across the maize genome responsible for dry weight are shown as black solid rectangular boxes. QTL hot regions for dry weight on chromosome were marked with a gray hollow rectangular box. (d) LOD curves of QTL mapping for SA and DW around the chromosome 7 QTLs at 6th time point. (e) LOD values of the bins at the peak of QTL interval are shown as a function of their genetic positions. (f) Candidate genes (GRMZM2G180490, GRMZM2G010702, GRMZM2G151649, and GRMZM2G057023) located in the peak bin. Other candidates genes located in the two bins next to the peak bin are listed in Supplementary Table 9.
Figure 5. Predication of maize digital biomass accumulation and yield. (a) Heat map showing the difference of digital biomass accumulation with different maize recombinant inbred lines. (b) The comparison of actual digital biomass (blue line) and predicated digital biomass (red line, using digital biomass of the first 6 time points to predicate the digital biomass of the rest 10 time points); error bars represent the standard error of dry weight of 167 samples in each time point. (c) The scatter plot shows the relationship between the actual grain yield and predicated yield with the predicted formula; a, b, c, d, e, f, g, h represent FDIC_1, LTA_above_1, GCV_8, SDLC_8, LTA_above_9, LTA_below_9, LNL_above_16, LSA_below_16, respectively; FDIC_1 and LNL_above_16 were leaf morphological traits; LTA_above_1, SDLC_8, LTA_above_9, LTA_below_9, and LSA_below_16 were leaf angle traits; GCV_8 was plant color traits; The black line was the fitting line and the standardized coefficients were shown in Supplementary Table 7; (d) The predicted ideotype maize plant based on the associated traits in early stages with higher GCV, FDIC, LNL_above, LTA_below, and lower LTA_above.
phenotyping. BMC Bioinformatics 12: 148


Copyright (c) 2020 American Society of Plant Biologists. All rights reserved.


