





# Phenomics-based GWAS analysis reveals the genetic architecture for drought resistance in cotton

Baoqi Li<sup>1</sup>, Lin Chen<sup>1</sup>, Weinan Sun<sup>1</sup>, Di Wu<sup>2,3</sup>, Maojun Wang<sup>1</sup>, Yu Yu<sup>4</sup>, Guoxing Chen<sup>5</sup>, Wanneng Yang<sup>1,2</sup> , Zhongxu Lin<sup>1</sup> , Xianlong Zhang<sup>1</sup> , Lingfeng Duan<sup>2,3,\*</sup> and Xiyan Yang<sup>1,\*</sup> 

<sup>1</sup>National Key Laboratory of Crop Genetic Improvement, National Center of Plant Gene Research (Wuhan), Huazhong Agricultural University, Wuhan, Hubei, China

<sup>2</sup>Hubei Key Laboratory of Agricultural Bioinformatics, Huazhong Agricultural University, Wuhan, Hubei, China

<sup>3</sup>College of Engineering, Huazhong Agricultural University, Wuhan, Hubei, China

<sup>4</sup>Cotton Institute, Xinjiang Academy of Agriculture and Reclamation Science, Shihezi, Xinjiang, China

<sup>5</sup>MOA Key Laboratory of Crop Ecophysiology and Farming System in the Middle Reaches of the Yangtze River, Huazhong Agricultural University, Wuhan, Hubei, China

Received 3 June 2019;

revised 13 February 2020;

accepted 5 June 2020.

\*Correspondence (Tel +86-27-87283955

(XY)/Tel +86-27-87282120 (LD); fax +86-

27-87280196 (XY)/fax +86-27-87287092

(LD); emails yxy@mail.hzau.edu.cn (XY) and

duanlingfeng@mail.hzau.edu.cn (LD))

## Summary

Drought resistance (DR) is a complex trait that is regulated by a variety of genes. Without comprehensive profiling of DR-related traits, the knowledge of the genetic architecture for DR in cotton remains limited. Thus, there is a need to bridge the gap between genomics and phenomics. In this study, an automatic phenotyping platform (APP) was systematically applied to examine 119 image-based digital traits (i-traits) during drought stress at the seedling stage, across a natural population of 200 representative upland cotton accessions. Some novel i-traits, as well as some traditional i-traits, were used to evaluate the DR in cotton. The phenomics data allowed us to identify 390 genetic loci by genome-wide association study (GWAS) using 56 morphological and 63 texture i-traits. DR-related genes, including *GhRD2*, *GhNAC4*, *GhHAT22* and *GhDREB2*, were identified as candidate genes by some digital traits. Further analysis of candidate genes showed that *Gh\_A04G0377* and *Gh\_A04G0378* functioned as negative regulators for cotton drought response. Based on the combined digital phenotyping, GWAS analysis and transcriptome data, we conclude that the phenomics dataset provides an excellent resource to characterize key genetic loci with an unprecedented resolution which can inform future genome-based breeding for improved DR in cotton.

**Keywords:** cotton, drought resistance, GWAS, i-traits, novel DR-related genes, phenomics.

## Introduction

Cotton (*Gossypium hirsutum* Linn.) is a relatively drought-resistant economic crop, and the fibre is a valuable source material for textiles (Chen *et al.*, 2007). Nevertheless, owing to the effects of climate change and extreme lack of fresh water (Li *et al.*, 2009a,b), drought has been one of the most critical abiotic stresses influencing the growth of crops (Zhu, 2002; Zhu, 2016). This leads to reduced yield in the field, even in cotton (Deeba *et al.*, 2012). Thus, there is an urgent need to improve drought resistance of cotton, not only through the use of irrigation but also through genetic improvement programmes.

Drought resistance (DR), a complex quantitative trait, is controlled by multiple interacting genes to induce morphological and physiological responses (Blum, 2011; Fukao and Xiong, 2013; Guo *et al.*, 2018; Zhou *et al.*, 2007). Plants respond differently at different growth stages under unpredictable hostile environments (Kooyers, 2015). In order to describe the complex mechanism of drought, an index defined as 'stress elasticity' has been used in barley, which shows a good correlation with drought tolerance (Chen *et al.*, 2014). In a previous study, natural variation in the gene *ZmVPP1* has been found to contribute to drought tolerance in maize seedlings, using survival rate as an indicator of drought tolerance (Wang *et al.*, 2016). Another study in rice led to the identification of *OsPP15* using a high-throughput automated phenotyping platform (Guo *et al.*, 2018). These studies support

the idea that plants might have their own phenotypic characters when subjected to water deficit, which would be useful to dissect drought resistance in cotton.

Phenomics represents a new interdisciplinary field that has attracted more attention in recent years, but still lags far behind genomics research (Houle *et al.*, 2010). Conventional phenotyping is labour-intensive, time-consuming, costly and low-throughput, and acquiring accurate phenotypes is one of the major bottlenecks in modern crop breeding (Furbank and Tester, 2011). However, high-throughput phenomics may avoid these drawbacks through the use of new technologies such as automatic phenotyping systems (Arvidsson *et al.*, 2011; Granier *et al.*, 2006; Nagel *et al.*, 2012; Walter *et al.*, 2007), noninvasive imaging (Al-Tamimi *et al.*, 2016; Araus and Cairns, 2014; Golzarian *et al.*, 2011; Yang *et al.*, 2014), and even artificial intelligence (Maded *et al.*, 2017; Potgieter *et al.*, 2017; Ubbens *et al.*, 2018). Multidimensional image data based on high-performance inputting have improved precision and saved time, and can provide scientists with more comprehensive information on plants. Massive plant phenomics datasets can be reanalysed by other users for integrating their omic data to verify and develop novel mechanisms and models (Cooper *et al.*, 2018; Fiorani and Schurr, 2013; Neveu *et al.*, 2019; Tardieu *et al.*, 2017). Integrated analysis of phenomics and genomics has contributed to the discovery of candidate genes in crops in recent years (Dingkuhn *et al.*, 2017; Prado *et al.*, 2018; Spindel *et al.*, 2018). Although

phenomics has been applied to many crop and fruit species (Aquino *et al.*, 2018; Sandmann *et al.*, 2018; Singh *et al.*, 2018; Yang *et al.*, 2015), its use in relation to drought stress has been less reported.

In previous work, an automatic phenotyping platform (APP) was utilized to achieve high-throughput screening of rice and maize for genetic studies and drought response (Guo *et al.*, 2018; Yang *et al.*, 2014; Zhang *et al.*, 2017). In the present study, the APP was expanded for high-throughput phenotyping of a natural cotton population consisting of 200 representative upland cotton accessions during the seedling stage of development once a week across six time points, either under standard watering or drought conditions. One hundred and nineteen different traits, with some traditional agronomic traits and more novel digital phenotypic traits, were obtained at each time point. Novel digital traits, as well as traditional DR-related traits, were identified as indicators for cotton DR. Combined with next-generation sequencing for genotyping, a total of 390 associated quantitative trait loci (QTLs), including some hotspots, were identified based on phenome and genome data. These QTLs included some previously reported genes: *GhRD2*, *GhNAC4*, *GhHAT22* and *GhDREB2* (Hou *et al.*, 2018; Li *et al.*, 2017), as well as novel genes.

## Results

### High-throughput automatic phenotyping platform for cotton

Phenotyping experiments were carried out in Wuhan using the APP (Yang *et al.*, 2014) with a natural population of 200 upland cotton accessions collected from 20 different regions (Fig. 1a; Table S1), grown either under standard watering or drought conditions (Fig. 1b). Each plant was inspected from the twenty side-view images at six time points with drought conditions at seedling stage (Fig. 1c), producing a total of 1 440 000 images, yielding ca. 1.5TB of data. After image processing, 119 image-based digital traits (i-traits) were collected at each time point. These could be divided into two categories – 56 morphological i-traits and 63 texture i-traits (Fig. 1d–e). These i-traits could be further used in GWAS (Fig. 1f–g).

The 56 morphological i-traits could be divided into four modules: plant geometry [plant height (PH), plant width (PW), height-width ratio (HWR), side total projected area – bounding rectangle area ratio (SBR), perimeter-area ratio (PAR), fractal dimension without and with image cropping (FD1, FD2)], plant compactness [plant density (PD); ( $i = 1, 2, 3, 4, 5$ , or 6) and PD derived features (PD<sub>1</sub>, PD<sub>2</sub>, PD<sub>3</sub>, PD<sub>R</sub>, PD<sub>K</sub> and PD<sub>S</sub>), relative frequencies (RF) <sub>$j$</sub>  ( $0 < j < 21$ ) ( $j = n$ ) and RF derived features (RF<sub>1</sub>, RF<sub>2</sub>, RF<sub>3</sub>, RF<sub>4</sub>, RF<sub>5</sub>, RF<sub>R1</sub>, RF<sub>R2</sub>, RF<sub>R3</sub>, RF<sub>11</sub>, RF<sub>21</sub>, RF<sub>31</sub>, RF<sub>R4</sub>, RF<sub>K</sub> and RF<sub>S</sub>)], plant colour [green area (GA), green projected area ratio (GPAR)] and plant biomass [side total projected area (SA)].

The 63 texture i-traits were computed from green (G), hue (H), and intensity (I) components of the RGB images. For each component, 21 texture i-traits were calculated, including 6 histogram traits [mean value (ME), standard deviation (SD), smoothness (ST), third moment (TM), uniformity (UF) and entropy (ET)] and 15 traits calculated from grey gradient co-occurrence matrix (GGCM<sub>1</sub>-GGCM<sub>15</sub>).

In order to evaluate the consistency of image-based measurement and manual measurement, plant height and fresh weight (biomass) were measured manually immediately after imaging. Strong positive correlations ( $R^2 > 0.93$ ,  $P$ -value  $< 0.001$  in plant

height, and  $R^2 > 0.93$ ,  $P$ -value  $< 0.001$  in fresh weight) were observed between image-based measurement and manual measurement (Fig. S1).

Phenotypes of cotton were affected by genotype and environment, especially affected by the latter at the seedling stage, when plants grew and changed rapidly every day, except for compactness traits: RF and middle-level PD (Table S2). However, the correlation between morphological i-traits revealed a similar trend in years 2015 and 2017 (Fig. S2a; Table S3), only some of texture i-traits were inconsistent in different years (Fig. S2b). A high positive correlation ( $r = 0.81$ ,  $P$ -value  $< 0.001$ , in year 2015;  $r = 0.77$ ,  $P$ -value  $< 0.001$ , in year 2017) was found between PH and PW, both i-traits being representative of plant geometry. Similarly, GA and GPAR showed an obvious correlation. For plant compactness, an obvious correlation was obtained between plant density PD <sub>$i$</sub>  and relative frequencies RF <sub>$j$</sub> . There were high positive correlations between low-level PD (PD<sub>1</sub> and PD<sub>2</sub>) and low-level RF (RF<sub>1</sub>, RF<sub>2</sub>, RF<sub>3</sub> and RF<sub>4</sub>), which were the important indices in showing the degree of the compactness at seedling stage for cotton. High correlations were also found for different types of i-traits, such as GA and SA (Fig. S2).

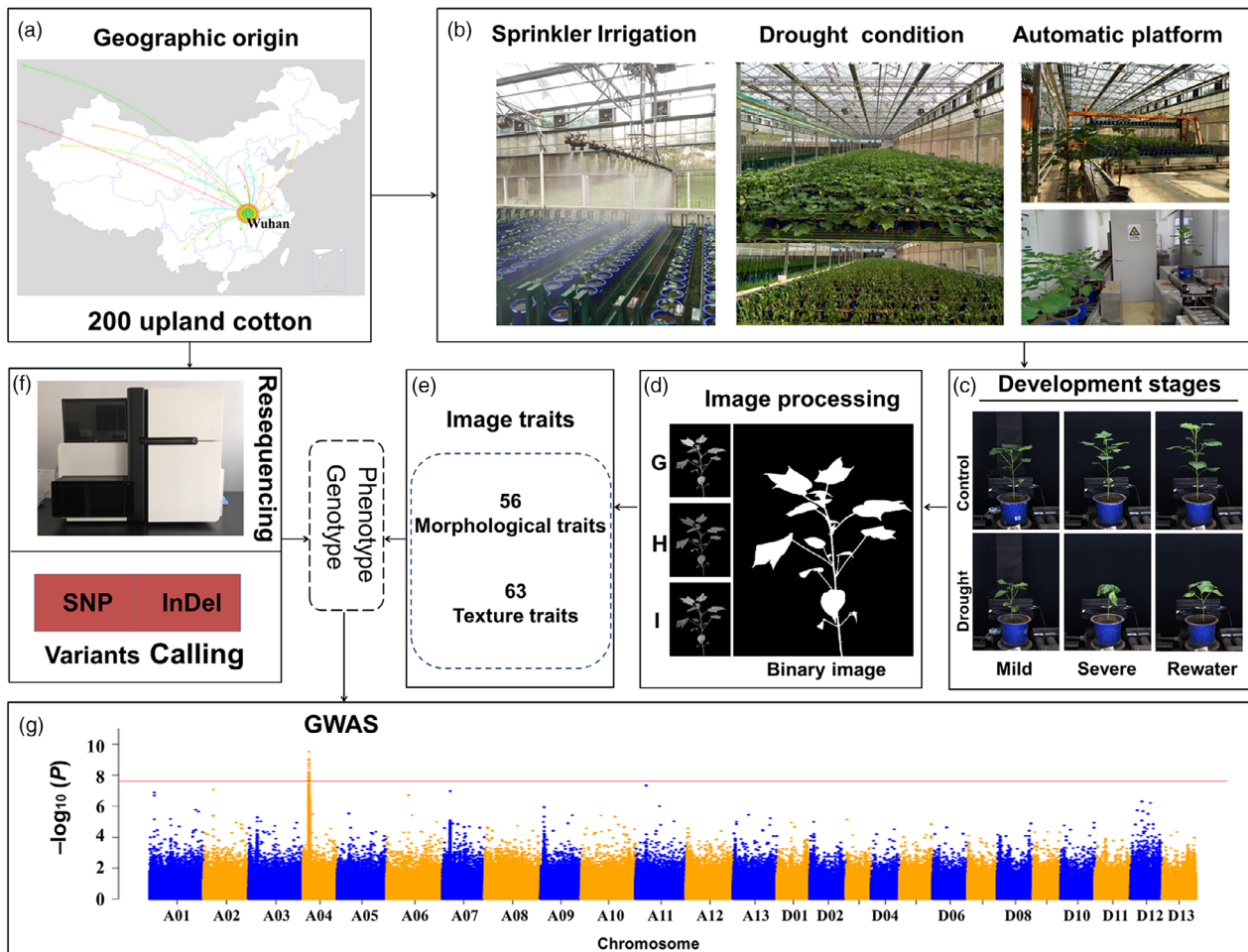
### Variation of i-traits under drought stress in cotton

To monitor the drought response of cotton during seedling stage (T<sub>1</sub>–T<sub>6</sub>) using the APP, we collected images of the 200 cotton accessions, withdrawing water from T<sub>2</sub> (mild drought) to T<sub>3</sub> (severe drought) time points, and then re-watering from T<sub>4</sub> (early-rewatering) to T<sub>6</sub> (late-rewatering). Plants of some accessions showed wilting, while plants of some accessions showed little reduced growth at T<sub>2</sub> and T<sub>3</sub>. These i-traits changed dramatically under drought conditions compared with under normal water condition, especially in geometric-related and compactness-related traits.

Coefficient of variation (CV) of drought-resistant coefficient (DRC) for different i-traits among 200 accessions was used to determine the discrete degree of observational data. The CV ranged from 0.01 to 1.78 across all morphological traits, showing large differences between different i-traits. By comparing the three sets of data with control, drought and DRC at six early growth stages, the CV became larger in the group of DRC, especially at severe drought stage (T<sub>3</sub>) for high-level compactness indicators (RF<sub>20</sub> and PD<sub>6</sub>) (Fig. S3).

PH and PW served as traditional indicators of drought response, which presented a significant difference between normal and drought groups. Taking two genotypes ZY7 (grouped as a drought-sensitive accession) and ZY116 (grouped as a drought-resistant accession) as examples, PH and PW reduced significantly under severe drought condition for ZY7 compared with ZY116, and an obvious symptom of wilting could be seen at T<sub>3</sub> under drought condition (Fig. 2a–c). After mild drought stress at T<sub>2</sub>, the increment of plant biomass was restricted and this trend continued with time and was maintained two weeks after re-watering (Fig. 2d). The GPAR showed little variation during seedling stage, but a large fluctuation was seen in compactness traits, such as RF<sub>20</sub> and PD<sub>6</sub>. In total, ninety morphological and texture i-traits showed a significant difference between treatment and control panels after drought stress (Paired  $t$ -test,  $P$ -value  $< 0.001$ ; Fig. S4) and most i-traits were in normal distribution.

Based on data from selected morphological i-traits (PD<sub>1–6</sub>, PAR, PW, PH, FD1, FD2, SBR, SA, HWR, GA and GPAR) under both



**Fig. 1** High-throughput automatic phenotyping platform for cotton. (a) Geographic origin of 200 collected upland cotton. (b) Drought experiments were carried out in the automatic phenotyping platform, plants in control were irrigated regularly by sprinkler facility. (c) Images were collected at six time points either under mild drought, severe drought and rewater treatment during seedling stage. (d) Image processing. Binary image and G, H, I components of plant image were extracted from RGB images. (e) Two types of *i*-traits were obtained: morphological traits and texture trait. (f-g) Combined with genome sequencing, these *i*-traits could be used in GWAS.

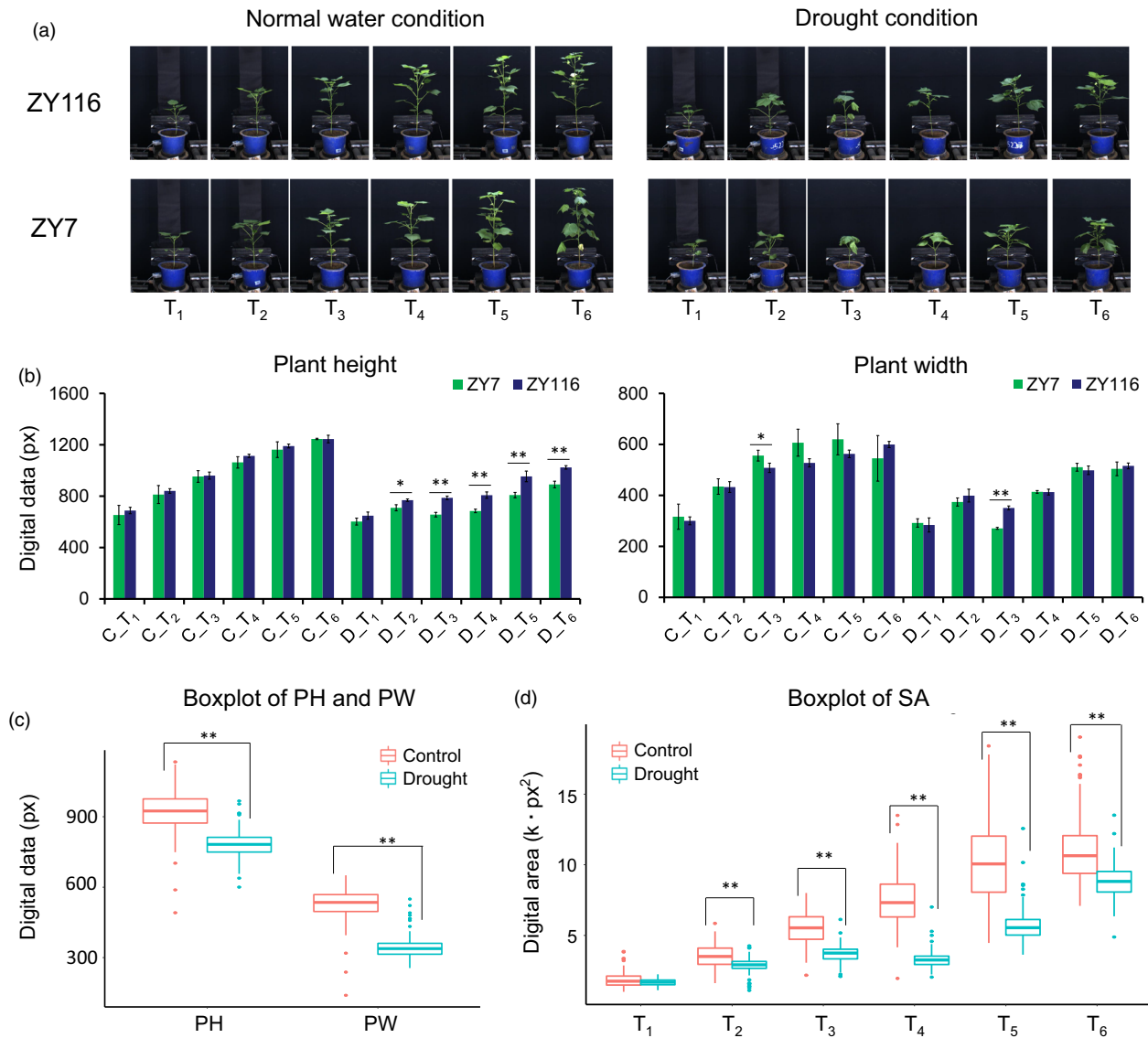
standard watering and drought conditions, the population of 200 upland cotton accessions could be divided into three sub-groups, whereby sub-groups I to III represented the cotton accessions that were affected severely or slightly by drought stress (Fig. S5). This result indicated that the vegetative growth of cotton plants was clearly restricted by water deficit, and variable responses to drought stress were observed among the population.

#### Novel *i*-traits could be used as indicators for DR

In addition to traditional agronomic traits, novel *i*-traits might be used to reflect the drought response. Compactness, which could be described by PD and RF, was an index that reflects the compactness of branches and leaves of plant. The plant density PD<sub>1-6</sub> stood for the ratio of six shading degrees, where PD<sub>1</sub> and PD<sub>6</sub> value represented the ratio of lowest and highest shading degree, respectively, and six features derived from PD [PD<sub>1</sub>, PD<sub>2</sub>, PD<sub>3</sub>, PD<sub>R</sub>, PD<sub>K</sub> and PD<sub>S</sub>] were used to give a comprehensive interpretation for drought response. Comparing two representative individuals, the distribution of PD<sub>1-6</sub> was basically the same of ZY116 under both drought and control conditions, but the distribution regularities of ZY7 were

completely different whereby the percentages of PD<sub>6</sub> and PD<sub>5</sub> were obviously higher under drought stress than control (Fig. 3a-b). Under severe water deficit, the treatment group showed an obviously higher PD<sub>6</sub> than control ( $P$ -value < 0.05) at T<sub>3</sub> (Fig. 3c), so that drought-sensitive and drought-resistant individuals could be identified easily using this index.

RF<sub>1-20</sub> represented twenty degrees of distribution for plants through twenty different angles, in which RF<sub>1</sub> and RF<sub>20</sub> represented the proportion of the lowest and highest compactness degree, respectively, and 14 traits derived from RF [RF<sub>1</sub>, RF<sub>2</sub>, RF<sub>3</sub>, RF<sub>4</sub>, RF<sub>5</sub>, RF<sub>R1</sub>, RF<sub>R2</sub>, RF<sub>R3</sub>, RF<sub>11</sub>, RF<sub>21</sub>, RF<sub>31</sub>, RF<sub>R4</sub>, RF<sub>K</sub> and RF<sub>S</sub>] were also used to evaluate the degree of drought response. RF<sub>R4</sub> was found to be effective in discriminating drought-resistant individuals and drought-sensitive individuals. For drought-sensitive individuals, they showed a higher RF<sub>R4</sub> value (RF<sub>R4</sub> > 0.2) than drought-resistant individuals (RF<sub>R4</sub> < 0.1) (Fig. 4a). To study the distribution of RF<sub>1-20</sub> of two representative plants of ZY7 and ZY116, a similar distribution was found for both individuals under normal water condition, but a large difference was observed in the distribution of ZY7 and ZY116 under drought conditions (Fig. 4b-c). These results



**Fig. 2** Variation of i-traits under drought stress during seedling stage. (a) Growth status images of two representative cotton plants (ZY7 and ZY116) at seedling stage (T<sub>1</sub>–T<sub>6</sub>) under normal water and drought conditions, respectively. (b) Statistical data for two geometric traits (plant height and plant width) of ZY7 and ZY116 at seedling stage. Prefix with C\_ and D\_ stand for control and drought treatment, respectively. (c–d) Statistical analysis for PW, PH and SA at group level. Error bars are calculated by standard deviation (s.d.) based on three biological replicates, statistical significance was determined by a two-sided *t*-test: \* *P*-value < 0.05, \*\* *P*-value < 0.01.

suggested that RF might be another useful i-trait to indicate drought response in cotton.

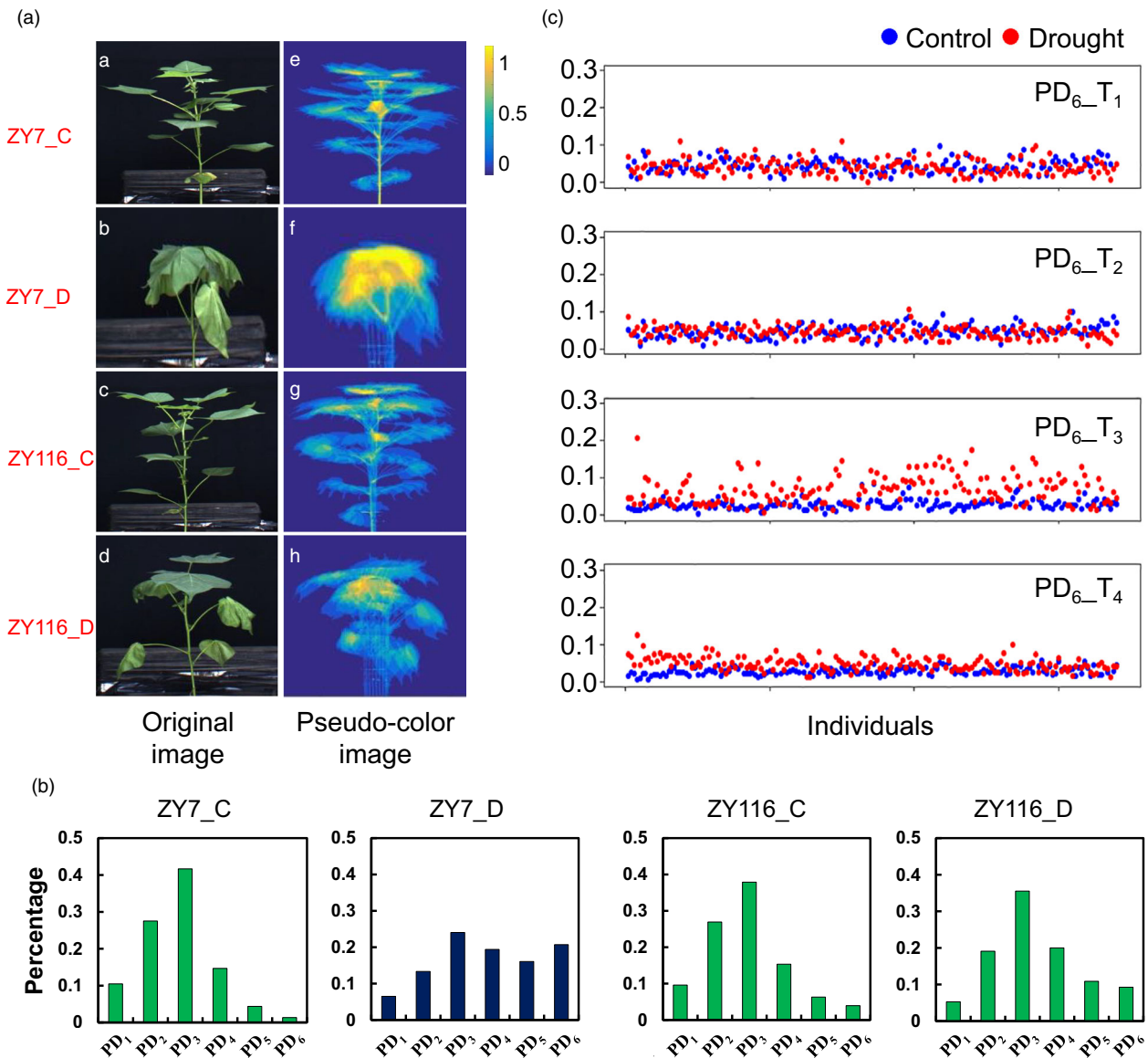
Some texture i-traits also could be used as indicators for drought response in cotton. Taking G as an example, G was the green component that to some extent reflects the chlorophyll content of plants. There were 21 traits in G texture. The DRC value of some texture i-traits, such as ET\_G, declined under drought conditions (Fig. 4d). Based on the above study, we take three representative accessions for further analysis. The results showed that the DRC value decreased under severe drought treatment to a lowest DRC value of ET\_G at 0.07 in the drought-sensitive accession ZY7, which suffered a more serious impact than two drought-resistant accessions ZY116 (DRC = 0.2) and ZY168 (DRC = 0.47). Moreover, the DRC value of ET\_G increased

to a relatively high-level after re-watering in the highly drought-resistant accession ZY168 (Fig. 4e).

These results show that dynamic growth is revealed by either traditional indicators or novel i-traits under drought treatments, providing a useful basis by which to find DR-related genes using GWAS.

#### Candidate DR-related genes based on GWAS and RNA-Seq

To analyse the molecular basis of DR-related i-traits, the cotton accessions were sequenced. A total of ca. 3.8 Tb of sequence data were integrated, with an average depth of ca. 7.4× (Table S1). These data were mapped against the TM-1 genome to identify genomic variants, and a total of 12 724 791 SNPs



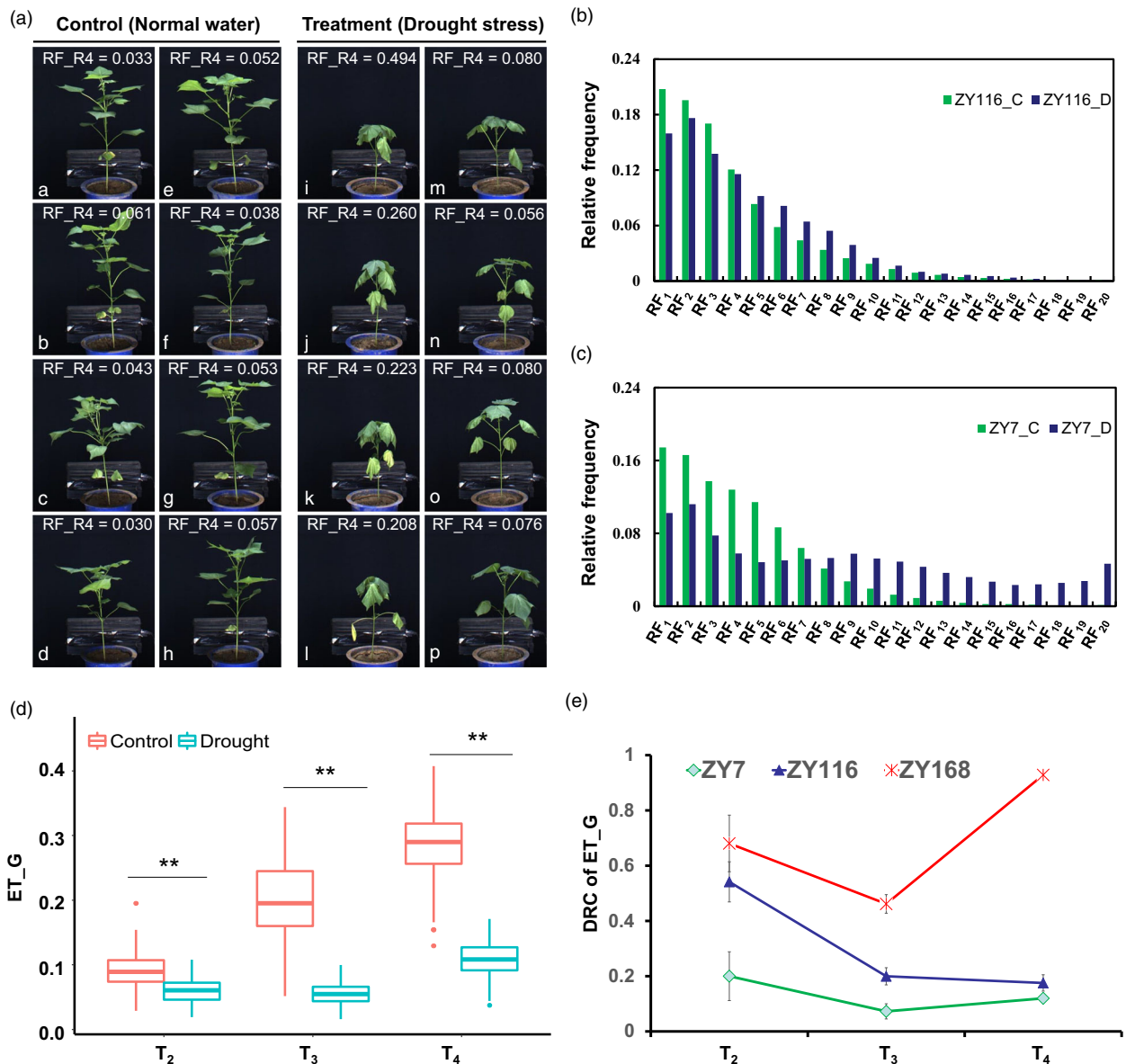
**Fig. 3** The plant compactness trait of plant density (PD). (a) Original and pseudo-colour images for cotton plants. **a-d** The original images of ZY7 and ZY116 under control and drought stress. **e-h** The right pictures illustrate the corresponding plant density images. (b) Plant density in different individuals. ZY7\_D has a different percentage distribution than ZY7\_C, while the distribution characteristics of ZY116\_C and ZY116\_D are similar. The suffix of \_C and \_D stands for control and drought treatment, respectively. (c) Scatter distribution map of indicator of PD<sub>6</sub> for cotton population at different seedling stages (T<sub>1-4</sub>) under control (blue filled dot) and drought stress (red filled dot).

were detected (Table S4). After filtering scaffold and MAF (minor allele frequency < 0.05) SNPs, 2 060 458 SNPs were used in factored spectrally transformed linear mixed models (FaST-LMM), and the strict significance *P*-value was set at  $2.43 \times 10^{-8}$  by Bonferroni correction. Population structure was calculated as the covariance to control the false positives (Fig. S6a). We explored the phylogenetic relationships among the 200 cotton accessions, using whole-genome SNP analysis, and those cotton accessions were divided into four groups, as supported by population structure analysis (Fig. S6b).

GWAS was performed across those cotton accessions for the value of RCRW (recovery capability after re-watering) of SA and DRC of 119 i-traits at different time points during seedling stage. A total of 622 unique SNPs were detected and with the smallest number of unique significant SNPs detected on Chromosome

(Chr) A05 (Table S5). Some loci identified by GWAS associated with some morphological and texture i-traits, as shown in Fig. S7a. Among them, the signal of one SNP peak in a region on Chr A04 showed gradual strengthening by GWAS analysis using SA at T<sub>2</sub> and T<sub>3</sub>, and finally exceeded the suggested threshold value at T<sub>4</sub> (Fig. S7b). Moreover, this SNP peak was found to be co-located by GWAS analysis using three i-traits (PH, PW and SA; Fig. 5a).

To find candidate genes around peak SNPs on different chromosomes, we needed to narrow down QTL regions. In our study, LD decay distance was ca. 500 kb using all SNPs (Fig. S8). A total of 390 candidate loci were detected by GWAS with those i-traits (Table S5), and 71 loci could be identified in both two years. Among them, 252 QTLs were associated with morphological i-traits, with a similar distributed amount to A and D

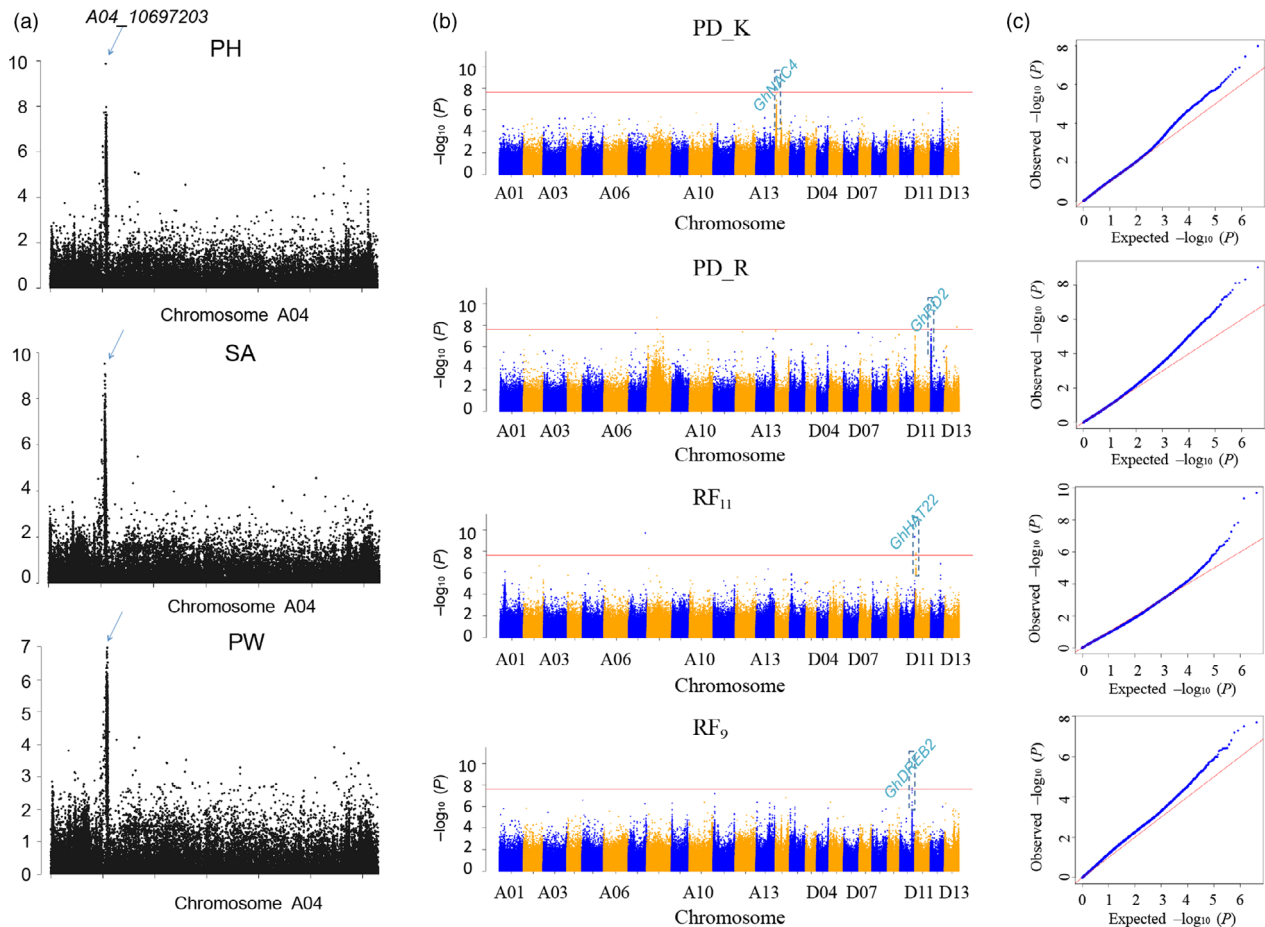


**Fig. 4** The plant compactness trait of relative frequency (RF) and texture i-traits (G). (a) RF\_R4 for sensitive and resistant cotton individuals. **a-h** ZY7, ZY283, ZY301, ZY303, ZY9, ZY116, ZY168 and ZY347 under normal water condition; **i-p** ZY7, ZY283, ZY301, ZY303, ZY9, ZY116, ZY168 and ZY347 under severe drought condition. (b) The distribution of RF<sub>1</sub> to RF<sub>20</sub> for a resistant plant of ZY116. (c) The distribution of RF<sub>1</sub> to RF<sub>20</sub> for a sensitive plant of ZY7. (d) Statistical analysis for ET\_G at group level. Error bars are calculated by standard deviation (s.d.) based on three biological replicates, statistical significance was determined by a two-sided *t*-test: \*\* *P*-value < 0.01. (e) The value of DRC of three representative cotton accessions (ZY7, ZY116 and ZY168). Error bars are calculated by standard deviation (s.d.) based on three biological replicates.

subgenome (132 versus 120; Fig. S9 and Table S5), although the size of the A subgenome (1477 Mb) is nearly twice that of the D subgenome (831 Mb) (Zhang *et al.*, 2015).

RNA-Seq was performed using a highly drought-sensitive cotton accession ZY7 and a relatively drought-resistant cotton accession ZY168 to differentially analyse the candidate genes at those associated loci. A total of 16 827 DEGs were found to be differentially expressed under drought treatments (*P*-adjust < 0.01, log<sub>2</sub>FoldChange > 1). Among them, 8074 DEGs were identified in both genotypes, with 2212 genes differentially expressed (*P*-value < 0.05) between the two. In addition, 5090 DEGs were identified only in ZY7, while 3663 DEGs were identified only in ZY168 (Table S6).

Fifteen QTLs within 167 DEGs through four i-traits of PD, RF, G and SA were selected for further analysis (Table S7). For these QTLs, QTL3 could be identified by GWAS with both RF and G. Four genes (*Gh\_D11G2874*, *Gh\_D11G2876*, *Gh\_D11G2884* and *Gh\_D11G2885*) were overlapping in QTL1 and QTL14 by GWAS with different i-traits. Some previously verified DR-related genes were co-localized within these regions. For example, *GhNAC4* and *GhRD2* were associated with the derived features of PD; and *GhDREB2* and *GhHAT22* were detected as the features of RF (Fig. 5b-c). However, more DEGs with homologous annotation of some reported stress-related genes in other plants were present at these loci, such as *Gh\_D11G2863* (ethylene responsive element binding factor 3), *Gh\_D13G2394* (tonoplast intrinsic



**Fig. 5** Genome-wide association studies. (a) Manhattan plots in Chr A04 for three traits of PH, SA and PW. The SNP that arrow points to as the same SNP described in the context. (b-c) Manhattan plots and QQ-plots for four representative novel i-traits. The QTLs including reported genes were marked by dotted boxes. Horizontal red full line indicates the genome-wide significant threshold ( $-\log_{10}P = 7.6$ ).

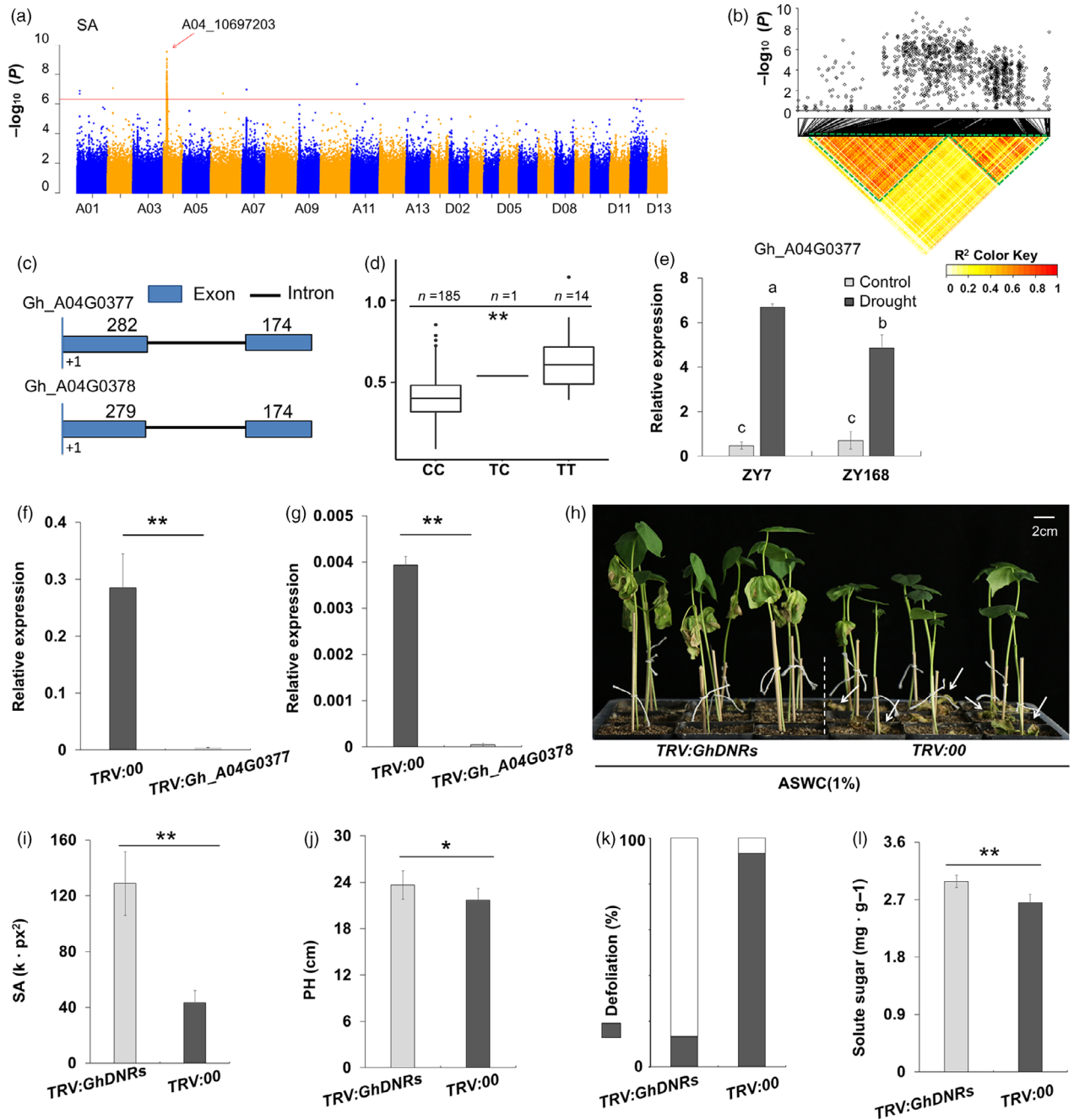
protein 1;3), *Gh\_D13G2393* (MYB domain protein 48), *Gh\_A12G1594* (WRKY DNA-binding protein 72), *Gh\_D11G2887* (transducin family protein/ WD-40 repeat family protein), *Gh\_D11G2899* (receptor-like protein 50) and *Gh\_D10G1405* (Calcium-binding EF-hand family protein; Table S7). Moreover, there were some DEGs without available annotation even for the homologous genes in *Arabidopsis* within the associated loci, such as *Gh\_D13G2320*, *Gh\_D13G2344*, *Gh\_D10G0952*, *Gh\_A04G0377*, *Gh\_A04G0378*, *Gh\_A12G1597*, *Gh\_A12G0141* and *Gh\_A12G0142*.

#### Identification of two novel DR-related negative regulative genes

As mentioned above, an associated locus on Chr A04 was identified by several i-traits, such as SA (Fig. 6a). Two closely linked sub-regions were presented in this region (Fig. 6b). Other than some annotated DEGs, two previously unreported tandem repeat genes, *Gh\_A04G0377* and *Gh\_A04G0378*, named as drought negative regulators (*GhDNRs*), were present in one sub-region (Table S7). These two genes each had two exons (Fig. 6c). The genotypes were either CC or TT alleles in the promoter regions of these two genes, and a significant difference in SA was found in two sets of plants – those that had the ‘TT’ allele show a higher biomass than the ‘CC’ variant under drought stress

(Fig. 6d). The relative expression level of *Gh\_A04G0377* showed a higher up-regulation in the drought-sensitive accession ZY7 than in the drought-resistant accession ZY168 under drought stress (Fig. 6e). These results suggested that this candidate gene could be transcriptionally induced by drought and repressed under normal water condition in the cotton.

Further functional validation of *GhDNRs* in cotton was carried out by virus-induced gene silencing (VIGS) technology. The drought stress was applied to the cotton plants at the seedling stage, with the two tandem repeat genes of *Gh\_A04G0377* and *Gh\_A04G0378* both silenced in the seedlings (Fig. 6f-h). The image-based biomass (SA) of *TRV:GhDNRs* plants was higher than the control *TRV:00* plants under drought stress (ASWC, ca. 1%; Fig. 6i), and the manually measured PH was higher in *TRV:GhDNRs* plants than the control *TRV:00* plants (Fig. 6j). The defoliation rate was decreased to 13.3% in the silenced plants compared with the 93.3% in control (Fig. 6k). Delayed leaf defoliation of *TRV:GhDNRs* plants was also observed in drought-sensitive plants (Fig. S10). *TRV:GhDNRs* plants accumulated more soluble sugar than control plants, which might contribute to drought tolerance (Fig. 6l). On the other hand, we found that silenced *GhDNRs* in the drought-resistant accession ZY168 showed only small differences in plant height and less difference in leaf defoliation in *TRV:GhDNRs* plants compared with control plants (Fig. S10).



**Fig. 6** Identification of the DR genes on Chr A04. (a) Manhattan plot of the trait of SA. The red line indicates the threshold of  $-\log_{10}P = 7.6$ , the red arrow points to the lead SNP of A04\_12697203. (b) LD heat map of surrounding the peak on Chr A04. Two interlocking regions within 472.3Kb are divided in the green dotted triangle. (c) Gene model of *Gh\_A04G0377* and *Gh\_A04G0378*. (d) Boxplots for SA, based on the genotypes of SNP A04\_10697203, which the red arrow points to in a. (e) Relative expression of *Gh\_A04G0377* in the leaf of ZY7 and ZY168 under control and drought stress. Values with different letters (a–c) indicate significant differences (Duncan's multiple comparisons,  $P$ -value < 0.05) between each sample. (f–g) qRT-PCR analysis confirmed that *Gh\_A04G0377* and *Gh\_A04G0378* were silenced in the *TRV:GhDNRs* plants. (h) VIGS of *GhDNRs* (*TRV:GhDNRs*) in drought-sensitive accession ZY7, and ZY7 with *TRV:00* was used as control. (i–j) Indicators of SA, manually measured PH, defoliation rate measured in control and silenced plants. (k) Soluble sugar content detected in leaves under drought stress. Error bars were calculated by S.D. based on three biological replicates, statistical significance was determined by a two-sided  $t$ -test: \*  $P$ -value < 0.05, \*\*  $P$ -value < 0.01.

## Discussion

The accurate identification and description of crop phenotypes are a prerequisite for a deep understanding of the relationship between phenotype and genotype as an important aim in

modern biology. There is also an urgent need to cultivate new breakthrough crop varieties and ensure national food production and food safety. Compared with the rapid development of genomic technology, the current methods for phenotyping are labour-intensive, costly, time-consuming and low-throughput, so



becoming a major bottleneck in crop improvement programmes (Chen *et al.*, 2014; Furbank and Tester, 2011). Here, we utilized efficient crop phenotyping techniques and acquired multidimensional i-traits for above-ground parts based on images generated by a high-throughput APP (Fig. 1), which has proved to be an efficient system in plant phenotyping (Guo *et al.*, 2018; Yang *et al.*, 2014; Zhang *et al.*, 2017).

Drought resistance is a complex quantitative trait. Plants have evolved drought adaptation in part via changes in morphological phenotype (Hu and Xiong, 2014). With the increasing amount of phenotypic data, phenomics has become a mode of investigation and several important phenotype databases are integrated for public analysis ([www.plantphenomics.com](http://www.plantphenomics.com); <http://www.plantphenomics.org.au/PODDProject>). The problem remains of how to filter out these redundant i-traits and pick out the data that meet our needs is another problem (Chen *et al.*, 2014; Schadt *et al.*, 2010). In our study, two kinds of i-traits were identified to study the drought response of cotton. Using the index of DRC, we studied the correlation among all morphological and texture i-traits to explain the change regularity of traits under drought conditions, which helped us to screen out some important candidate traits, such as PH, SA, PD and RF.

Plant height and some other traditional traits are vital indices for cotton DR (Fig. 2). In the current study, we paid more attention to novel image-based indicators (PD, RF and G) to explain the DR of cotton in different dimensions. Drought-sensitive individuals presented a higher percentage of PD<sub>6</sub> under drought stress (Fig. 3). RF<sub>1</sub> to RF<sub>20</sub>, and their derived feature from RF could also be used to evaluate DR and allow visualization of the drought-resistant individuals and drought-sensitive individuals (Fig. 4a-c). One trait in texture i-traits was also selected to display the DR response (Fig. 4d-e).

We performed GWAS on a population of 200 upland cotton accessions with the phenotypes collected from the above APP for DR during seedling stage. We not only co-located some previously reported DR-related genes in cotton (Fig. 5), but also found some proposed meaningful stress-related genes, such as *Gh\_D13G2393*, and *Gh\_A12G1594* (Table S7). Moreover, due to the high positive correlation between PH, SA and PW, a QTL was co-located on Chr A04. Within this region, two previously unreported tandem repeat genes, *Gh\_A04G0377* and *Gh\_A04G0378*, were found to relate to drought stress.

To the best of our knowledge, this is the first study to conduct GWAS based on high-throughput APP derived multiple i-traits in a cotton drought study. During seedling growth, dynamic DR was revealed in cotton, which might be influenced by different genes. Among a large population, drought-sensitive and drought-resistant individuals presented different growth models. Yield and quality of fibre were not investigated in this study, which are the two important traits for the value of cotton. Accurate phenotypic identification data for field-grown plants was hard to obtain because of exposure to a range of biotic and abiotic stresses. Nevertheless, we associated high-density resequencing information with abundant phenotypic data and identified novel genes related to DR, as well as previously reported relevant genes (Table S7). Some candidate genes were verified as being differentially expressed under drought stress, and novel genes, *GhDNRs* were shown to negatively regulate drought tolerance at physiological and transcriptional levels through VIGS experiments, or gene expression tests that showed the drought-sensitive cotton accession exhibited more significant up-regulation of relative expression of the genes (Fig. 6). The *GhDNRs*

genes might control the biomass and defoliation under drought stress, but the deeper mechanisms of drought tolerance remained to be clarified.

## Materials and methods

### Plant materials and experimental design

A natural population of 200 representative upland cotton accessions was planted in APP facilities (Yang *et al.*, 2014) at the National Key Laboratory of Crop Genetic Improvement in Huazhong Agricultural University, Wuhan, Hubei Province, P.R.China (30°28'27"N, 114°21'1"E, 30 m elevation) with two year replicates and three replicates each year, either under standard watering or drought conditions during seedling stage. Among them, 186 cotton accessions were collected from the major Chinese cotton cultivation regions: the northwestern inland region (NIR), the northern specific early maturation region (NSEMR), the Yellow River region (YRR) and the Yangtze River region (YtRR), another 14 varieties were introduced from America and Soviet Union (AS), as described previously (Huang *et al.*, 2017; Wang *et al.*, 2017). Information on the cotton accessions is provided in Table S1. Seeds were sown in late April (Table S8). Plants were grown under controlled greenhouse conditions and were watered by an automatic irrigation system at germination stage, and then manually watered with about 900 mL/pot once a week three days before images capturing. Fertilization was carried out at sowing, and a compound boron fertilizer was added before T<sub>5</sub> (60 kg of water + 100 g of boric fertilizer + 370 g of carbamide, fully dissolved and 150 mL applied per pot).

Cotton plants were subject to automatic phenotyping at the three-leaf stage (ie 30 days after sowing in year 2015, and 32 days after sowing in year 2017). All the cotton individuals were screened at a one-week interval to generate six time points (as T<sub>1</sub> to T<sub>6</sub>) during the seedling stage. Drought stress was conducted by withdrawing water from the plants (as D) at T<sub>2</sub> (mild drought) to T<sub>3</sub> (severe drought) time points, while control plants were well watered. Absolute soil water content (ASWC) was used to evaluate the degree of drought in cotton [the ASWC of mild drought (T<sub>2</sub>) was ca. 5%, and 1% for severe drought (T<sub>3</sub>), while ca. 15% for the parallel control]. The experiments were conducted on three biological replicates. Each plant grown in a pot with 4.5 kg soil was used as a biological replicate. In total, 1200 plants were used and imaged in this study. The line interval size was 400 mm, and individual plant interval size was 250 mm with one line remaining empty between two planted lines. A small population of ten representative varieties was planted for manual measurement and digital modelling. The sowing dates, watering dates and detection dates for two years were provided in Table S8.

### Phenotyping and image-based digital traits (i-traits) extraction

The details of the phenotyping facility were described previously (Yang *et al.*, 2014). The automatic phenotyping of cotton was conducted as shown in Fig. 1. RGB images of the cotton plants were acquired using a modified high-throughput rice phenotyping facility (HRPF). Twenty side-view images (2452 × 2056 pixels) were taken from different angles at one time point for each plant. Image processing and features extraction were accomplished using LabVIEW as described previously, with minor modifications (Yang *et al.*, 2014). Manual measurements were conducted at

each time-point for digital modelling. After image processing, 119 image-based digital traits (i-traits) were computed, including 56 morphological i-traits and 63 texture i-traits.

The textures calculated from green (G), hue (H), and intensity (I) component were denoted as G texture, H texture and I texture. For instance, mean value in G texture was denoted as ME\_G. The detailed definition and equation of all the i-traits were shown in the File S1.

### Analysis of DR-related i-traits

CV using value of control (suffix with \_C), drought (suffix with \_D) among 200 accessions for different i-traits was used to determine the discrete degree of observational data. We focused on mild drought condition ( $T_2$ ), severe drought condition ( $T_3$ ) and early-rewatering treatment ( $T_4$ ). The changes observed in some i-traits on these three time points were expected to reflect the drought response of cotton. These dynamic changes were investigated in two ways: (i) i-trait changes following mild to severe drought conditions (such as traditional indicator, PH and some novel i-traits as defined below); (ii) i-trait changes representing phenotypic recovery following serious water deficiency (such as SA). DRC (ratio between i-trait<sub>drought</sub> and i-trait<sub>control</sub>) and RCRW [(i-trait<sub>T<sub>4</sub></sub> - i-trait<sub>T<sub>3</sub></sub>)/i-trait<sub>T<sub>3</sub></sub>] were used to evaluate the phenotypic response. Analysis of variance (ANOVA) across two years was performed using QTL IciMapping software (Meng *et al.*, 2015).

The DRC of all traits at  $T_2$ ,  $T_3$  and  $T_4$  was used to analyse the performance of the accessions under mild drought, severe drought and rewatering conditions. RCRW was defined to show the recovery capability by re-watering after drought treatments. The values of DRC (suffix with \_R) were also used to analyse the correlations among the morphological i-traits and texture i-traits, respectively. The distribution of all i-traits was tested by Shapiro-Wilk test.

### Genome-wide association studies for DR-related i-traits

Of the 200 cotton accessions, 170 were sequenced in a previous study (Wang *et al.*, 2017) and another 30 accessions were sequenced in 2017 (Table S1). The allotetraploid cotton genome (*Gossypium hirsutum* L. acc. TM-1) was used as reference genome (Zhang *et al.*, 2015). Clean paired-end reads (PE150) were mapped to the TM-1 genome using BWA software (Li and Durbin, 2009) with the default parameters. Picard-tools in Genome Analysis ToolKit (GATK) (McKenna *et al.*, 2010) software were used to sort BAM format files, remove duplicated data and realign InDel reads. SNPs and InDels calling were calculated separately by two different softwares: GATK and SAMtools (Li *et al.*, 2009a,b), and linkage disequilibrium (LD) for all pairs of SNPs was calculated using PopLDdecay with the default parameters (Zhang *et al.*, 2019).

GWAS was performed with FaST-LMM (version 2.02) (Lippert *et al.*, 2011). The RCRW of SA and DRC of all i-traits was used for GWAS. All filtered SNPs used in association analysis were compressed into binary formats using PLINK software (version 1.90b6.8) (Purcell *et al.*, 2007). The variant annotations were annotated using the software SnpEff (version 4.3) (Cingolani *et al.*, 2012). All SNPs were categorized as being in intergenic regions, upstream and downstream regions, in exons or introns by default parameters. The significant association threshold values were set at  $-\log_{10}(0.05/n)$  ( $n$ , total filtered SNP number). The significant association regions were manually verified from the aligned resequencing reads against the TM-1 genome with SAMtools.

### Differential expression analysis and validation of candidate genes

Leaves of the accession of ZY7 (drought-sensitive) and ZY168 (drought-resistant) from treated and parallel control plants were sampled with two biological replicates for RNA-Seq analysis at severe drought stage ( $T_3$ ) just as for phenotyping. Total RNA was extracted using the TG-DP441 RNA Kit, according to the manufacturer's instructions. A total of 2  $\mu$ g RNA were used for library construction, and RNA sequencing was performed on an Illumina Novaseq™ 6000 system. Clean reads (PE150) were mapped to the TM-1 genome with Hisat2 (version 2.1.0) (Kim *et al.*, 2015). The expression level of each gene was determined by StringTie (v1.3.3b) (Pertea *et al.*, 2015), and the differentially expressed genes (DEGs) ( $P$ -adjust < 0.01,  $\log_2$ FoldChange > 1) were calculated by DESeq2 R package.  $t$ -Test was implemented to check the significance between the changes of transcript abundance between the two cotton accessions under drought conditions. qRT-PCR was conducted to verify the expression pattern of candidate genes.

VIGS was used to validate the candidate genes as described previously (Gao *et al.*, 2011). The coding region for *Gh\_A040377* was cloned into the TRV2 (tobacco rattle virus) vector. The constructed vectors were transformed into *A.tumefaciens* GV3101 by electroporation and agroinfiltrated into the cotyledons of drought-sensitive accession ZY7 and drought-resistant accession ZY168 through syringe inoculation when its two cotyledons had fully spread (13 days after sowing). qRT-PCR was conducted with the ABI 7500 system (Applied Biosystems, Foster City, CA) to verify the expression of the candidate genes, with the internal control *GhUBQ7*. The  $2^{-\Delta\Delta CT}$  method was used to present relative changes in gene expression levels (Schmittgen and Livak, 2008). Primers are listed in Table S9. Drought stress was applied to the plants by withdrawing water after seed germination in water-saturated soil. After plants germinated with no watering for 30 days (ASWC, ca. 1%), the phenotypes were collected. Leaves were sampled to verify the silencing effect when the second leaf had fully emerged. Nondestructive biomass (SA) was extracted just as for phenotyping. The plant height and defoliation of leaves were measured manually. Soluble sugar was determined by Anthrone-Sulfuric acid colorimetry (Yemm and Willis, 1954). Significant test was performed by  $t$ -test ( $P$ -value < 0.05).

### Acknowledgements

Funding was provided by the National Key Project of Research and Development Plan (2018YFD1000907), National Natural Science Foundation of China (31560410), and the Fundamental Research Funds for the Central Universities (2662017PY058, 2662020ZKPY011).

### Conflict of interest

The authors have declared that no conflict of interest exists.

### Author contributions

X.Y. and X.Z. conceived the project and designed the experiments; B.L. conducted the phenotyping and GWAS experiments; W.S. and D.W. participated in the phenotyping; W.Y. participated in designing the APP platform; G.C. contributed to the cotton

cultivation; Z.L. and Y.Y. provided the seed materials; L.D. extracted imaging traits from raw images; L.C. conducted the VIGS experiments; M.W. offered writing advice; B.L. wrote the manuscript and managed the main data analysis, including statistical analysis, figure and table design; X.Y. revised the manuscript.

### Data Availability Statement

All the data sets generated during the current study are available in the NCBI Sequence Read Archive (SRA) under project number PRJNA530048. The accession numbers sequenced in this study are summarized in Table S1.

### References

- Al-Tamimi, N., Brien, C., Oakey, H., Berger, B., Saade, S., Ho, Y.S., Schmockel, S.M. *et al.* (2016) Salinity tolerance loci revealed in rice using high-throughput non-invasive phenotyping. *Nat. Commun.* **7**, 13342.
- Aquino, A., Barrio, I., Diago, M.-P., Millan, B. and Tardaguila, J. (2018) vitisBerry: An Android-smartphone application to early evaluate the number of grapevine berries by means of image analysis. *Comput. Electron. Arg.* **148**, 19–28.
- Araus, J.L. and Cairns, J.E. (2014) Field high-throughput phenotyping: the new crop breeding frontier. *Trends Plant Sci.* **19**, 52–61.
- Avidsson, S., Perez-Rodriguez, P. and Mueller-Roeber, B. (2011) A growth phenotyping pipeline for *Arabidopsis thaliana* integrating image analysis and rosette area modeling for robust quantification of genotype effects. *New Phytol.* **191**, 895–907.
- Blum, A. (2011) Drought resistance – is it really a complex trait? *Funct. Plant Biol.* **38**, 753–757.
- Chen, Z.J., Scheffler, B.E., Dennis, E., Triplett, B.A., Zhang, T., Guo, W., Chen, X. *et al.* (2007) Toward sequencing cotton (*Gossypium*) genomes. *Plant Physiol.* **145**, 1303–1310.
- Chen, D., Neumann, K., Friedel, S., Kilian, B., Chen, M., Altmann, T. and Klukas, C. (2014) Dissecting the phenotypic components of crop plant growth and drought responses based on high-throughput image analysis. *Plant Cell*, **26**, 4636–4655.
- Cingolani, P., Patel, V.M., Coon, M., Nguyen, T., Land, S.J., Ruden, D.M. and Lu, X. (2012) Using *Drosophila melanogaster* as a model for genotoxic chemical mutational studies with a new program. *SnpSift. Front. Genet.* **3**, 35.
- Cooper, L., Meier, A., Laporte, M.-A., Elser, J.L., Mungall, C., Sinn, B.T., Cavaliere, D. *et al.* (2018) The Planteome database: an integrated resource for reference ontologies, plant genomics and phenomics. *Nucleic Acids Res.* **46**, D1168–D1180.
- Deeba, F., Pandey, A.K., Ranjan, S., Mishra, A., Singh, R., Sharma, Y.K., Shirke, P.A. *et al.* (2012) Physiological and proteomic responses of cotton (*Gossypium herbaceum* L.) to drought stress. *Plant Physiol. Biochem.* **53**, 6–18.
- Dingkuhn, M., Pasco, R., Pasuquin, J.M., Damo, J., Soulie, J.C., Raboin, L.M., Dusserre, J. *et al.* (2017) Crop-model assisted phenomics and genome-wide association study for climate adaptation of indica rice. 2. Thermal stress and spikelet sterility. *J. Exp. Bot.* **68**, 4389–4406.
- Fiorani, F. and Schurr, U. (2013) Future scenarios for plant phenotyping. *Annu. Rev. Plant Biol.* **64**, 267–291.
- Fukao, T. and Xiong, L. (2013) Genetic mechanisms conferring adaptation to submergence and drought in rice: simple or complex? *Curr. Opin. Plant Biol.* **16**, 196–204.
- Furbank, R.T. and Tester, M. (2011) Phenomics—technologies to relieve the phenotyping bottleneck. *Trends Plant Sci.* **16**, 635–644.
- Gao, X., Britt, R.C. Jr, Shan, L. and He, P. (2011) Agrobacterium-mediated virus-induced gene silencing assay in cotton. *J. Vis. Exp.* **54**, 1745.
- Golzarian, M.R., Frick, R.A., Rajendran, K., Berger, B., Roy, S., Tester, M. and Lun, D.S. (2011) Accurate inference of shoot biomass from high-throughput images of cereal plants. *Plant Methods.* **7**, 2.
- Granier, C., Aguirrezabal, L., Chenu, K., Cookson, S.J., Dauzat, M., Hamard, P., Thioux, J.J. *et al.* (2006) PHENOPSIS, an automated platform for reproducible phenotyping of plant responses to soil water deficit in *Arabidopsis thaliana* permitted the identification of an accession with low sensitivity to soil water deficit. *New Phytol.* **169**, 623–635.
- Guo, Z., Yang, W., Chang, Y., Ma, X., Tu, H., Xiong, F., Jiang, N. *et al.* (2018) Genome-wide association studies of image traits reveal the genetic architecture of drought resistance in rice. *Mol. Plant.* **11**, 789–805.
- Hou, S., Zhu, G., Li, Y., Li, W., Fu, J., Niu, E., Li, L. *et al.* (2018) Genome-wide association studies reveal genetic variation and candidate genes of drought stress related traits in cotton (*Gossypium hirsutum* L.). *Front. Plant Sci.* **9**, 1276.
- Houle, D., Govindaraju, D.R. and Omholt, S. (2010) Phenomics: the next challenge. *Nat. Rev. Genet.* **11**, 855–866.
- Hu, H. and Xiong, L. (2014) Genetic engineering and breeding of drought-resistant crops. *Annu. Rev. Plant Biol.* **65**, 715–741.
- Huang, C., Nie, X., Shen, C., You, C., Li, W., Zhao, W., Zhang, X. *et al.* (2017) Population structure and genetic basis of the agronomic traits of upland cotton in China revealed by a genome-wide association study using high-density SNPs. *Plant. Biotechnol. J.* **15**, 1374–1386.
- Kim, D., Langmead, B. and Salzberg, S.L. (2015) HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods.* **12**, 357–360.
- Kooyers, N.J. (2015) The evolution of drought escape and avoidance in natural herbaceous populations. *Plant Sci.* **234**, 155–162.
- Li, H. and Durbin, R. (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, **25**, 1754–1760.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G. *et al.* (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, **25**, 2078–2079.
- Li, Y., Ye, W., Wang, M. and Yan, X. (2009) Climate change and drought: a risk assessment of crop-yield impacts. *Clim. Res.* **39**, 31–46.
- Li, F., Li, M., Wang, P., Cox, K.L. Jr, Duan, L., Dever, J.K., Shan, L. *et al.* (2017) Regulation of cotton (*Gossypium hirsutum*) drought responses by mitogen-activated protein (MAP) kinase cascade-mediated phosphorylation of GhWRKY59. *New Phytol.* **215**, 1462–1475.
- Lippert, C., Listgarten, J., Liu, Y., Kadie, C.M., Davidson, R.I. and Heckerman, D. (2011) FaST linear mixed models for genome-wide association studies. *Nat. Methods*, **8**, 833–835.
- Madec, S., Baret, F., de Solan, B., Thomas, S., Dutartre, D., Jezequel, S., Hemmerle, M. *et al.* (2017) High-throughput phenotyping of plant height: comparing unmanned aerial vehicles and ground LiDAR estimates. *Front. Plant Sci.* **8**, 2002.
- McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernysky, A., Garimella, K. *et al.* (2010) The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303.
- Meng, L., Li, H., Zhang, L. and Wang, J. (2015) QTL IciMapping: integrated software for genetic linkage map construction and quantitative trait locus mapping in biparental populations. *Crop J.* **3**, 269–283.
- Nagel, K.A., Putz, A., Gilmer, F., Heinz, K., Fischbach, A., Pfeifer, J., Faget, M. *et al.* (2012) GROWSCREEN-Rhizo is a novel phenotyping robot enabling simultaneous measurements of root and shoot growth for plants grown in soil-filled rhizotrons. *Funct. Plant Biol.* **39**, 891.
- Neveu, P., Tireau, A., Hilgert, N., Negre, V., Mineau-Cesari, J., Bricchet, N., Chapuis, R. *et al.* (2019) Dealing with multi-source and multi-scale information in plant phenomics: the ontology-driven Phenotyping Hybrid Information System. *New Phytol.* **221**, 588–601.
- Pertea, M., Pertea, G.M., Antonescu, C.M., Chang, T.C., Mendell, J.T. and Salzberg, S.L. (2015) StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* **33**, 290–295.
- Potgieter, A.B., George-Jaeggli, B., Chapman, S.C., Laws, K., Suarez Cadavid, L.A., Wixted, J., Watson, J. *et al.* (2017) Multi-spectral imaging from an unmanned aerial vehicle enables the assessment of seasonal leaf area dynamics of sorghum breeding lines. *Front. Plant Sci.* **8**, 1532.
- Prado, S.A., Cabrera-Bosquet, L., Grau, A., Coupel-Ledru, A., Millet, E.J., Welcker, C. and Tardieu, F. (2018) Phenomics allows identification of genomic regions affecting maize stomatal conductance with conditional

- effects of water deficit and evaporative demand. *Plant Cell Environ.* **41**, 314–326.
- Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J. et al. (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575.
- Sandmann, M., Grosch, R. and Graefe, J. (2018) The use of features from fluorescence, thermography, and NDVI imaging to detect biotic stress in Lettuce. *Plant Dis.* **102**, 1101–1107.
- Schadt, E.E., Linderman, M.D., Sorenson, J., Lee, L. and Nolan, G.P. (2010) Computational solutions to large-scale data management and analysis. *Nat. Rev. Genet.* **11**, 647–657.
- Schmittgen T.D., Livak K.J. (2008) Analyzing real-time PCR data by the comparative  $C_T$  method. *Nature Protocols*, **3** (6), 1101–1108. <http://dx.doi.org/10.1038/nprot.2008.73>
- Singh, A.K., Ganapathysubramanian, B., Sarkar, S. and Singh, A. (2018) Deep learning for plant stress phenotyping: trends and future perspectives. *Trends Plant Sci.* **23**, 883–898.
- Spindel, J.E., Dahlberg, J., Colgan, M., Hollingsworth, J., Sievert, J., Staggenborg, S.H., Huttmacher, R. et al. (2018) Association mapping by aerial drone reveals 213 genetic associations for Sorghum bicolor biomass traits under drought. *BMC Genom.*, **19**, 679.
- Tardieu, F., Cabrera-Bosquet, L., Pridmore, T. and Bennett, M. (2017) Plant phenomics, from sensors to knowledge. *Curr. Biol.* **27**, R770–R783.
- Ubbens, J., Cieslak, M., Prusinkiewicz, P. and Stavness, I. (2018) The use of plant models in deep learning: an application to leaf counting in rosette plants. *Plant Methods*, **14**, 6.
- Walter, A., Scharr, H., Gilmer, F., Zierer, R., Nagel, K.A., Ernst, M., Wiese, A. et al. (2007) Dynamics of seedling growth acclimation towards altered light conditions can be quantified via GROWSCREEN: a setup and procedure designed for rapid optical phenotyping of different plant species. *New Phytol.* **174**, 447–455.
- Wang, X., Wang, H., Liu, S., Ferjani, A., Li, J., Yan, J., Yang, X. et al. (2016) Genetic variation in ZmVPP1 contributes to drought tolerance in maize seedlings. *Nat. Genet.* **48**, 1233–1241.
- Wang, M., Tu, L., Lin, M., Lin, Z., Wang, P., Yang, Q., Ye, Z. et al. (2017) Asymmetric subgenome selection and cis-regulatory divergence during cotton domestication. *Nat. Genet.* **49**, 579–587.
- Yang, W., Guo, Z., Huang, C., Duan, L., Chen, G., Jiang, N., Fang, W. et al. (2014) Combining high-throughput phenotyping and genome-wide association studies to reveal natural genetic variation in rice. *Nat. Commun.* **5**, 5087.
- Yang, W., Guo, Z., Huang, C., Wang, K., Jiang, N., Feng, H., Chen, G. et al. (2015) Genome-wide association study of rice (*Oryza sativa* L.) leaf traits with a high-throughput leaf scorer. *J. Exp. Bot.* **66**, 5605–5615.
- Yemm, E.W. and Willis, A.J. (1954) The estimation of carbohydrates in plant extracts by anthrone. *Biochem. J.* **57**, 508–514.
- Zhang, T., Hu, Y., Jiang, W., Fang, L., Guan, X., Chen, J., Zhang, J. et al. (2015) Sequencing of allotetraploid cotton (*Gossypium hirsutum* L. acc. TM-1) provides a resource for fiber improvement. *Nat. Biotechnol.* **33**, 531–537.
- Zhang, X., Huang, C., Wu, D., Qiao, F., Li, W., Duan, L., Wang, K. et al. (2017) High-throughput phenotyping and QTL mapping reveals the genetic architecture of maize plant growth. *Plant Physiol.* **173**, 1554–1564.
- Zhang, C., Dong, S.S., Xu, J.Y., He, W.M. and Yang, T.L. (2019) PopLDdecay: a fast and effective tool for linkage disequilibrium decay analysis based on variant call format files. *Bioinformatics*, **35**, 1786–1788.
- Zhou, J., Wang, X., Jiao, Y., Qin, Y., Liu, X., He, K., Chen, C. et al. (2007) Global genome expression analysis of rice in response to drought and high-salinity stresses in shoot, flag leaf, and panicle. *Plant Mol. Biol.* **63**, 591–608.
- Zhu, J.K. (2002) Salt and drought stress signal transduction in plants. *Annu. Rev. Plant Biol.* **53**, 247–273.
- Zhu, J.K. (2016) Abiotic stress signaling and responses in plants. *Cell*, **167**, 313–324.

## Supporting information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

**Figure S1.** Correlation between manual measurement and image-based measurement in trait of plant height and biomass.

**Figure S2.** Correlation analysis for morphological and texture i-traits in two years, respectively.

**Figure S3.** Coefficient of variation (CV) among 200 accessions for three sets of at seedling stage.

**Figure S4.** T-test between treatment and control panels at seedling stage.

**Figure S5.** Three sub-groups i-traits of 200 upland accessions.

**Figure S6.** Structure analysis of 200 upland cotton accessions.

**Figure S7.** GWAS results by morphological and texture i-traits.

**Figure S8.** Linkage disequilibrium (LD) decay distance at whole genome level.

**Figure S9.** Distribution of QTLs associated with morphological traits in 26 chromosomes in cotton.

**Figure S10.** VIGS of GhDNRs (TRV:GhDNRs) in drought-resistant accession ZY168 under drought stress.

**Table S1.** Cotton resources information and resequencing information used in this study.

**Table S2.** ANOVA for morphological traits across years in 2015 and 2017.

**Table S3.** Correlation analysis of morphological traits over two years.

**Table S4.** Variants rate details.

**Table S5.** Information of associated loci.

**Table S6.** DEGs in two cotton accessions under drought stress.

**Table S7.** QTLs with DEGs in four i-traits.

**Table S8.** Date of two-year seedling experiment of cotton.

**Table S9.** List primers used in this study.

**SupFile S1.** The image processing and i-traits extraction.