Deciphering genetic basis of developmental and agronomic traits by integrating high-throughput optical phenotyping and genome-wide association studies in wheat

Jie Gao[1,†], Xin Hu[1,†], Chunyan Gao[1], Guang Chen[1], Hui Feng[1], Zhen Jia[1], Peimin Zhao[1], Haiyang Yu[1], Huaiwen Li[1], Zedong Geng[1], Jingbo Fu[1], Jun Zhang[1], Yikeng Cheng[1], Bo Yang[1], Zhanghan Pang[1], Daoquan Xiang[2] (iD), Jizeng Jia[3] (iD), Handong Su[1,5] (iD), Hailiang Mao[1], Caixia Lan[1,5] (iD), Wei Chen[1,5] (iD), Wenhao Yan[1,5,*] (iD), Lifeng Gao[3,*], Wanneng Yang[1,5,*] (iD) and Qiang Li[1,4,*] (iD)

[1]National Key Laboratory of Crop Genetic Improvement, Huazhong Agricultural University, Wuhan, China

[2]Aquatic and Crop Resource Development, National Research Council Canada, Saskatoon, Saskatchewan, Canada

[3]Institute of Crop Sciences, Chinese Academy of Crop Sciences (CAAS), Beijing, China

[4]The Center of Crop Nanobiotechnology, Huazhong Agricultural University, Wuhan, China

[5]Hubei Hongshan Laboratory, Huazhong Agricultural University, Wuhan, China

Summary

Dissecting the genetic basis of complex traits such as dynamic growth and yield potential is a major challenge in crops. Monitoring the growth throughout growing season in a large wheat population to uncover the temporal genetic controls for plant growth and yield-related traits has so far not been explored. In this study, a diverse wheat panel composed of 288 lines was monitored by a non-invasive and high-throughput phenotyping platform to collect growth traits from seedling to grain filling stage and their relationship with yield-related traits was further explored. Whole genome re-sequencing of the panel provided 12.64 million markers for a high-resolution genome-wide association analysis using 190 image-based traits and 17 agronomic traits. A total of 8327 marker-trait associations were detected and clustered into 1605 quantitative trait loci (QTLs) including a number of known genes or QTLs. We identified 277 pleiotropic QTLs controlling multiple traits at different growth stages which revealed temporal dynamics of QTLs action on plant development and yield production in wheat. A candidate gene related to plant growth that was detected by image traits was further validated. Particularly, our study demonstrated that the yield-related traits are largely predictable using models developed based on i-traits and provide possibility for high-throughput early selection, thus to accelerate breeding process. Our study explored the genetic architecture of growth and yield-related traits by combining high-throughput phenotyping and genotyping, which further unravelled the complex and stage-specific contributions of genetic loci to optimize growth and yield in wheat.

Introduction

Allohexaploid bread wheat (*Triticum aestivum L.*, $2n = 6x = 42$, AABBDD) is one of the most important crops around the world, providing more than 20% of the calories and protein consumed by humans (FAO, 2020). Although significant progress has been made in wheat breeding programs to improve important agronomic traits (Gao et al., 2017), enormous efforts are still needed to fill the gap between population growth and wheat production (Tester and Langridge, 2010). A number of studies in crops indicated that yield is affected by early plant growth and biomass accumulation (DeWitt et al., 2021; Gowda et al., 2011; Nadolska-Orczyk et al., 2017). Therefore, understanding the genetic basis for wheat growth and yield potential is of significant importance for breeding practice.

In the last decades, genes associated with specific traits such as plant height (*Rht1* and *Rht8*), vernalization (*Vrn1*, *Vrn2*, *Vrn3*), spikelets number (*FRIZZY PANICLE*) and spike architecture (*TaCOL-B5*) have been identified and functionally studied through positional cloning in wheat (Dobrovolskaya et al., 2015; Pearce et al., 2011; Xiong et al., 2022; Yan et al., 2003, 2006; Zhang et al., 2022). More recently, a number of QTLs were identified through genome-wide association studies (GWAS) with limited markers and a few easy-to-measure traits collected at a particular growth stage (Cao et al., 2020; Jamil et al., 2019; Pang et al., 2020). While sequencing technologies have provided almost unlimited high-density genetic markers, rapid and accurate phenotyping in a large-scale manner remains a major constraint for genetic studies (Yang et al., 2014). High-throughput phenotyping platform (HTP), equipped with non-

invasive technologies, offers an opportunity to capture growth traits at multiple developmental stages to address the phenotyping bottleneck and uncover temporal genetic controls underlying plant development or stress adaptation (Watt *et al*., 2020). The integration of genomics and phenomics has been applied to dissect the genetic basis of complex traits in rice (Yang *et al*., 2014), cotton (Li *et al*., 2020a), rapeseed (Li *et al*., 2020b) and maize (Wu *et al*., 2021; Zhang *et al*., 2017). In wheat, a study monitored 12 growth traits of 208 recombinant inbred lines (RILs) through HTP, only found a few QTLs with major effects due to low marker density and small population size (Camargo *et al*., 2018).

Plant growth is a dynamic and complicated process controlled by several genes/QTLs and determines biomass and yield traits (Knoch *et al*., 2020; Li *et al*., 2020b; Würschum *et al*., 2014; Zhang *et al*., 2017). Non-invasive and imaging-based phenotyping platform provides the opportunity to link plant growth with yield. A number of marker-trait associations and QTLs for biomass and growth-related traits were detected through GWAS using data collected from multiple time points during early growth of rapeseed seedlings (Knoch *et al*., 2020; Li *et al*., 2020b). So far, most studies in wheat have focused on either yield traits or growth traits at a particular developmental stage (Pang *et al*., 2020; Watt *et al*., 2020). However, the yield potential and plant growth dynamics are both complex traits controlled by numerous genes/QTLs, and most of them only hold small or medium effects. Our knowledge of what these factors are and how they are genetically and temporally controlled in a stage-specific manner is largely unknown.

In this study, we performed whole genome re-sequencing of 288 elite wheat lines representing a wide genetic and geographical diversity and collected the dynamic growth traits of each accession from seedling to mature stage using the high-throughput phenotyping platform. We conducted GWAS and investigated the relationship between image-based traits (i-traits) and agronomic traits. 52 pleiotropic QTLs affecting both i-traits and agronomic traits were detected, suggesting a strong link between those two classes of traits. A number of growth-related candidate genes were identified in the i-trait QTLs including known genes involved in plant growth and development. A novel gene involved in regulating plant growth and yield traits was further validated through mutant analyses. Prediction models derived using i-traits from earlier growth stages largely increased the prediction accuracy for final yield traits. This study highlights the importance of i-traits in capturing temporal growth patterns to unravel the complex genetics of growth and yield in wheat.

Results

Genetic diversity and population structure of wheat germplasm collection

In this study, a total of 288 wheat lines with wide genomic diversity were selected for whole genome re-sequencing (Data S1). In order to obtain high-density markers, we generated 11.19 Tb of high-quality reads from these accessions. A total of 12.18 million single nucleotide polymorphisms (SNPs) and 0.46 million insertion–deletion (InDel) markers were obtained (Data S2). There were 5.13, 6.34 and 1.17 million markers on A, B and D subgenome (Figure 1a; Data S2). An average of 601 858 polymorphic markers was identified per chromosome, and these markers were enriched towards the chromosome ends. Among these markers, 163 362 (1.29%) were located in gene body, 17 909 (0.14%) were located
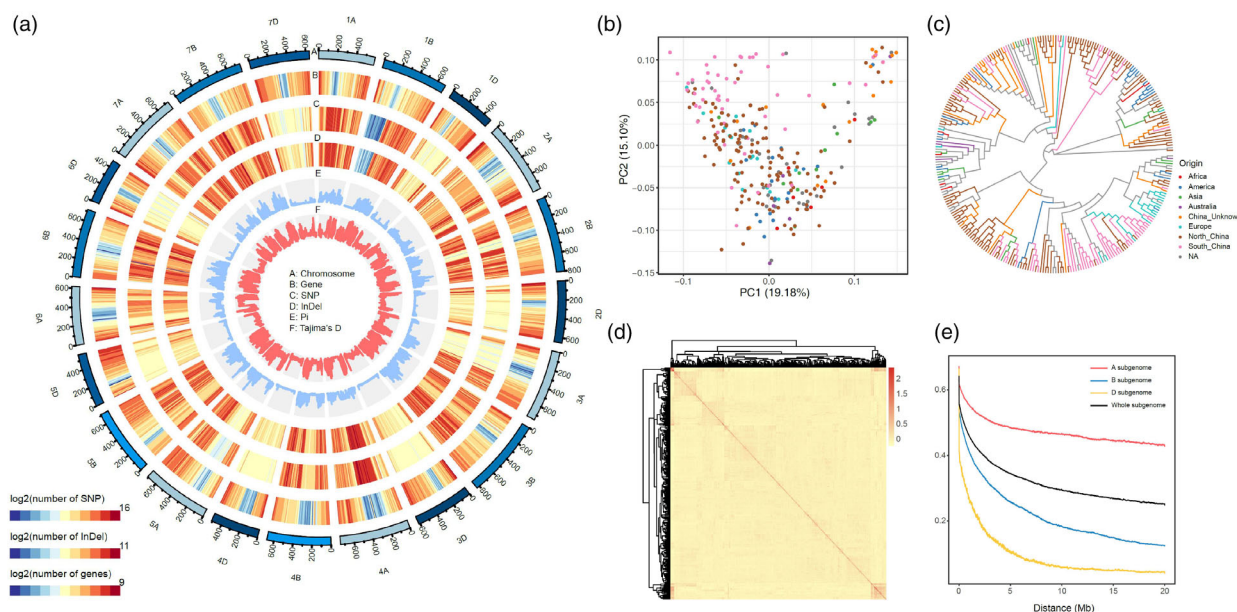
Figure 1 Whole genome re-sequencing, genetic diversity and population structure of the wheat association panel. (a) Circular diagram showing genomic diversity of the association panel. The tracks on the plot represent different genomic features: A, Chromosome name and size; B, Density of gene models; C, Density of SNP markers; D, Density of InDels; E, Average nucleotide diversity (Pi) at variable sites within a sliding window of 1 Mb; F, Neutrality test statistics (Tajima's D) at variable sites within a sliding window of 1 Mb. (b) Principal component analysis of 288 accessions. Proportions of explained variance of principal components (PCs) 1 and 2 are indicated on the axes. Different colours represent geographical origin of wheat lines. (c) Neighbour-jointing tree analysis for the association panel. (d) Heatmap of pairwise kinship matrix. (e) Linkage disequilibrium (LD) decay for the whole genome and A, B and D subgenomes. Decay curves of A, B, D and whole genome were presented in squared correlation of allele frequencies at diallelic loci (r^2).
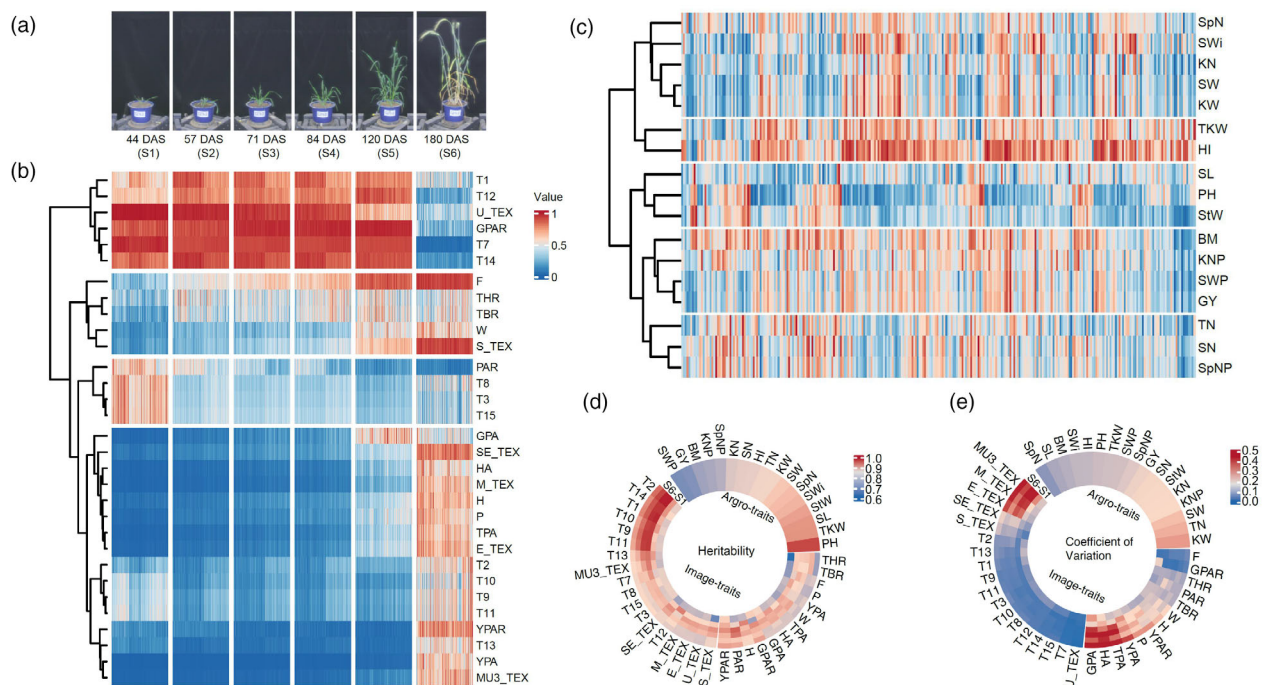
Figure 2 Phenotype diversity of the association panel. (a) Representative side-view images of wheat at six growth stages. Each individual plant was monitored using optical cameras at 44 days after sown (DAS) (S1, seedling), 57 DAS (S2, early tillering), 71 DAS (S3, late tillering), 84 DAS (S4, stem elongation), 120 DAS (S5, heading) and 180 DAS (S6, grain filling). (b) Clustering analysis of 31 image traits (i-traits) captured at six growth stages. Normalized values of i-traits were clustered into four distinct groups. (c) Heatmap of 17 agronomic traits. The agronomic traits were clustered into five groups. (d) Heatmap of broad-sense heritability (H^2) for 186 i-traits and 17 agronomic traits. The i-traits were presented from S6 to S1 towards the centre of circle. (e) Heatmap of coefficient of variation (CV) for 186 i-traits and 17 agronomic traits. The i-traits were presented from S6 to S1 towards the centre of circle.

in untranslated region, and the remaining 12 469 974 (98.66%) were located in intergenic regions (Data S2). Notably, 0.27% of markers located in exon led to 34 527 non-synonymous mutations.

Principal component analysis (PCA) showed moderate variation with PC1 (19.18%), PC2 (15.10%) and PC3 (13.40%). Biplots showed that accessions from different geographical origin were apart from each other and landraces were generally separated from modern varieties (Figures 1b and S1). The neighbour joining tree generated similar results (Figure 1c). In consistent, majority of wheat lines were unrelated to each other (Figure 1d). Admixture analysis showed the cross-validation error dropped dramatically when K value increased from 1 to 9 and then stabilized (Figure S2). Thus, these results suggested a high genetic diversity of the population.

All markers were used to calculate genome-wide mean linkage disequilibrium (LD). The LD statistic r^2 was used to estimate LD in sliding window size of 20 Mb for each of the 21 chromosomes (Figure 1e). Consistent with previous study (Pang *et al.*, 2020), the faster LD decays in B and D subgenome compared to A subgenome might be related to their independent evolutionary history and different selection pressure during domestication and modern breeding among three subgenomes.

High-throughput phenotyping of the wheat association panel

To obtain the dynamic growth traits and yield-related traits, the wheat accessions were cultivated under greenhouse conditions with three biological replicates for each line (Figures 2a and S3). Employing an automated and high-throughput phenotyping

platform (HTP) (Yang *et al.*, 2014), the dynamic growth pattern of each individual plant was monitored using optical cameras at six growing stages covering seedling (S1), early tillering (S2), late tillering (S3), stem elongation (S4), heading (S5) and grain filling (S6) stages. Six side-view pictures were taken for each plant at different development stage, generating 31 104 images in total. A total of 78 morphological and 108 texture image-based traits (i-traits) were extracted from the side-view pictures (Figure 2b; Data S3). Four growth parameters were derived from exponential models for H (height of the bounding rectangle) and TPA (total projected area; Data S4). Moreover, 17 yield-related agronomic traits were measured, including spike length (SL), kernel number per spike (KN), thousand kernel weight (TKW), biomass, etc (Figure 2c; Data S5).

Image traits changed dramatically along the plant development and were grouped into four clusters according to their dynamic patterns (Figure 2b). Six i-traits including five texture traits for grey and gradient homogeneity of pixels (T1, T12, T7, T14, U_TEX) and green projected area ratio (GPAR) were at high levels from seedling to heading stages (S1-S5) but dropped dramatically at the last stage which coincides with the leaf senescence at maturation. On the contrary, four i-traits including PAR (the ratio of perimeter and total projected area), T8, T3 and T15 only showed higher levels at S1, and steady decreased from S2 to S6. The values of five i-traits including F (fraction dimension), THR (the ratio of total projected area and hull area), TBR (the ratio of total projected area and circumscribed box area) and W (width of the bounding rectangle) gradually increased along plant development. The largest cluster has 16 i-traits including GPA (green

projected area), TPA (total projected area) and H (height of the bounding rectangle) which are good indicators of plant growth. These results solidly supported that i-traits could reflect the variable growth pattern of wheat.

To compare the phenotypic variation and heritability among the accessions, we assessed coefficient of variation (CV) and broad-sense heritability (H^2) for 186 i-traits and 17 agronomic traits (Figure 2d,e). The 186 i-traits were composed of 31 i-traits measured at each of the six growth stages (S1 to S6). In general, the morphological i-traits such as TPA, HA (hull area) and GPA were more variable. About 56% of i-traits and 41% of agronomic traits showed a H^2 greater than 0.85 (Data S6). The yield component traits, including spike weight per plant (SWP), grain yield (GY), biomass (BM), kernel number per plant (KNP) and spikelets number per plant (SpNP), showed lower level of H^2, which suggested that they were complex traits controlled by numerous loci with small effect. In general, i-traits showed higher level of variation and heritability, which could be more preferable in GWAS analysis.

Consistent with previous studies in other crops (Knoch et al., 2020; Li et al., 2020b; Muraya et al., 2017; Yang et al., 2014; Zhang et al., 2017), we observed significant correlations between i-traits and agronomic traits including PH (plant height), StW (straw weight) and other yield component traits (Figure S4; Data S7). Straw weight (StW) was highly correlated with 5 i-traits including plant size indicators, TPA_S6 and HA_S6 ($R > 0.6$; P-value <0.01). Biomass (BM) was also significantly correlated with TPA_S6. Different from StW and BM, which were correlated with i-traits detected at S6, tiller number (TN) was highly related with i-traits detected at early stages such as indicators for plant compactness, THR_S2, TBR_S2 and TBR_S3. In summary, 78 i-traits were significantly correlated ($R > 0.3$ or < -0.3, P-value-<0.01) with 12 yield-related traits (Data S7).

Genome-wide association analysis and candidate gene identification

To elucidate the genetic basis of phenotype variation, we performed genome-wide association analysis using 207 traits and 12.64 million markers. A total of 8327 marker-trait associations (MTAs) were detected for 174 traits (P-value <1E-7) (Figure 3a; Data S8). These markers were clustered into 1605 QTLs by defined LD blocks ($r^2 < 0.5$). In total, 136 QTLs for agronomic traits and 1469 QTLs for i-traits were detected (Figure 3a,b), of which, 384 QTLs were co-localized with previously reported QTLs in wheat (Data S9). Notably, 346 unique QTLs for 118 i-traits collected at during plant development were co-localized with reported QTLs for yield, thousand kernel weight (TKW), grain weight (GW), etc. A number of known genes (e.g. *Vrn1-5A*, *Rht8*, *Ppd-D1*) were identified in these QTLs (Figure 3a; Data S9).

The 1605 QTLs were unevenly distributed among the A, B and D subgenomes, with 567 on the A subgenome, 687 on the B subgenome and 351 on the D subgenome (Data S8). The QTLs for each agronomic trait ranged from one for TN (tiller number) to 35 for PH (plant height), whereas this number ranged from one for F_S6 to 79 for YPAR_S2 for i-traits (Figure 3c; Data S8). Importantly, 1085 (67.6%) of 1605 QTLs spanned on intervals less than 1 Mb (Figure S5). 33 QTLs were identified with each explained more than 20% phenotypic variation (PVE), including 4 QTLs for agronomic traits and 29 QTLs for image traits (Data S8). More QTLs were detected for i-traits than agronomic traits, especially for those with larger PVE, suggesting that i-traits are more stable and effective for GWAS.

A total of 6856 candidate genes were identified in the QTLs regions, including 1317 genes for agronomic traits and 5880 genes for i-traits (Data S10). Candidate genes were further annotated by taking their rice orthologs as reference (Kawahara et al., 2013). A number of them were involved in meristem development and hormone metabolism, such as *Vrn1-5A* (*TraesCS5A03G0935400*), *Vrn3-7A* (*TraesCS7A03G0272100*), *Vrn1-5D* (*TraesCS5D03G0894800*), *TaAGL11* (*TraesCS6B03G0 969600*) (Data S10). SNP annotation showed that 36 genes harbour 50 non-synonymous mutation or nonsense mutations. These genes were selected as candidate genes with particular interests (Data S11). For example, one SNP (P-value = 8.62E-07) causing missense mutation was identified in *TraesCS6B03G010 5200* for P_S6 (perimeter of plant at stage S6), an indicator of plant size and biomass (Data S11). *TraesCS6B03G0105200* encodes a high-affinity nitrate transporter (NRT) (O'Brien et al., 2016; Ohkubo et al., 2021; Shi et al., 2022), thus suggesting its involvement in nitrogen uptake and biomass accumulation. Although further investigation and validation of candidate genes is required, this study indicated that the time-series characteristics of i-traits depicting the growth dynamics allowed for functional mapping of dynamic QTLs underlying plant development.

Temporal dynamics of QTLs action on plant growth and yield in wheat

A number of genetic studies have found evidence for QTLs with pleiotropic effects which affects multiple traits (Cui et al., 2016; Fan et al., 2019). We investigated the overlapping regions of QTLs detected for individual traits and identified a total of 277 pleiotropic QTLs including 7 for agronomic traits, 218 for image traits and 52 for both agronomic and image traits (Data S12). A trait-QTLs network was constructed, and two hub QTLs, *QTL168* and *QTL171*, were revealed (Figure S6). *QTL168* was detected for 50 i-traits while *QTL171* was detected for 41 i-traits, suggesting a major genetic loci controlling multiple traits. We further investigated the pleiotropic QTLs associated with both agronomic and image traits (Figure S7). For example, three QTLs were commonly detected for StW (straw weight) and P_S6 (perimeter at S6) which was consistent with the high correlation (R = 0.62) between the two traits. The pleiotropic QTLs for both agronomic and image traits thus approved the linkage between the two types of traits.

A total of 171, 259, 243, 206, 347 and 174 QTLs were identified at S1, S2, S3, S4, S5 and S6 stage, respectively (Data S8). We specifically investigated the dynamic changes of QTLs for 31 i-traits at multiple stages (Figure 3b). Different number and genomic position of QTLs were detected for each of the 31 i-traits at the six growth stages (Figure 3b,d; Data S8), indicating the dynamics of QTLs in regulating the developmental behaviour of these quantitative traits. Of the 1469 image QTLs, 260, 90 and 45 QTLs were detected at two, three or four stages, respectively. Interestingly, 126 of them were located at three positions (*QTL168*, *QTL170* and *QTL171*) on chromosome 5D (Figure S6), which were also classified as pleiotropic QTLs based on their association with multiple traits. *QTL168* is associated with 50 i-traits at S1 to S5 stages and a candidate gene *Vrn1-5D* (also called *Vrn-D1*), which promotes heading and flowering dates in wheat (Zhang et al., 2012, 2018), is located in this region. In addition, pleiotropic *QTL139*, which contains *Vrn1-5A*, contributes to TBR_S3, T1_S4 and T2_S4. We genotyped *Vrn1-5A*, *Vrn1-5B* and *Vrn1-5D* in the association population. Consistent with GWAS

Figure 3 Dynamic QTLs for i-traits and agronomic traits detected by GWAS. (a) Manhattan plots for all traits measured in this study. Upper panel showed the results from 17 agronomic traits, and lower panel represents the results from 157 i-traits. The QTLs derived by significantly associated markers were present as sticks in the middle. QTLs for agronomic traits were shown in brown colour, and QTLs for i-traits were displayed in blue. 52 QTLs (shown in black colour) were commonly identified by two classes of traits. (b) Dynamics of QTLs detected by individual traits at each of the six stages. Yellow dots, QTLs for agronomic traits; blue dots, QTLs for i-traits; red dots, pleiotropic QTLs.(c) Number of QTLs detected for each of the individual traits. (d) Number of QTLs detected at multiple growth stages. QTLs detected at one, two, three and four stages were shown in the stacked barplot as blue, brown, green and purple, respectively.
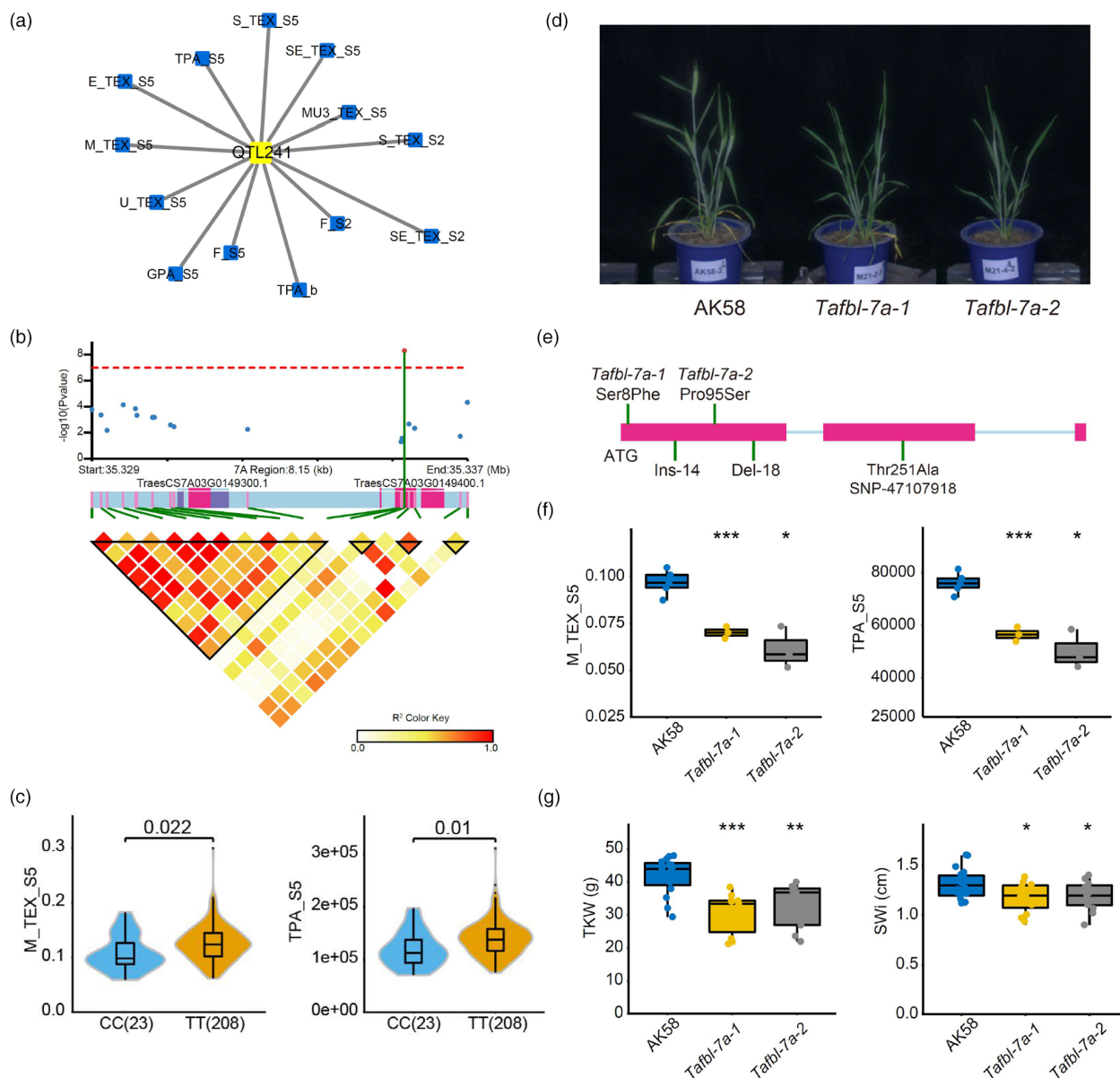
Figure 4 Validation of candidate gene for *QTL241* on chromosome 7A. (a) *QTL241* is associated with 13 i-traits. (b) Manhattan plot of the *QTL241* on chromosome 7A. Zoom in region (from 35.32 to 35.34 Mb) was shown and the non-synonymous marker SNP-47107918 was significantly associated with i-trait TPA_S5 (*P*-value = 4.85E-09). The SNP is located on the second exon of *TraesCS7A03G0149400* as shown in the gene model over the LD heatmap. (c) Plants with the CC allele of SNP-47107918 showed significantly lower levels of TPA_S5 and F_S5 than those with the TT allele. (d) The side-view image of AK58 and two individual mutants at heading stage (S5). (e) Point mutations were detected on first exon of *TraesCS7A03G0149400* and conformed by Sanger sequencing. (f) Mutants showed significantly lower levels of M_TEX_S5 and TPA_S5 than AK58. (g) Mutants showed significantly lower levels of TKW and SWi than AK58.

analysis, our results showed that lines harbouring dominant alleles of *Vrn1-5D* and *Vrn1-5A* flower earlier than those with recessive alleles (Figure S8; Data S13). Together, these results demonstrated that the genetic basis of growth dynamics can be dissected and pleiotropic QTLs are not only for static trait but also in regulating dynamic plant development at multiple stages.

Variation in *TraesCS7A03G0149400* affects both i-traits and yield traits

QTL241 was commonly detected for 13 i-traits including TPA_S5, F_S5 and GPA_S5 (Figure 4a). The leading SNP (SNP-47107918, C/T) was located exactly in the second exon of *TraesCS7A03G0149400*, causing a missense mutation (Figure 4b). Protein sequence analysis showed that *TraesCS7A03G0149400* encodes a cyclin-like F-box protein (Figure S9; Hong *et al.*, 2020; Xu *et al.*, 2009). Wheat lines with CC allele showed significantly lower levels of TPA_S5 and M_TEX_S5 when compared to lines with TT allele (Figure 4c). Besides the i-traits, the lines with CC haplotype also showed reduced SWi (spike width) and TKW (thousand kernel weight) in comparison with lines harbouring TT haplotype (Figure S10). Two individual EMS mutant lines possessing point mutations causing Ser8Phe and Pro95Ser were

obtained and confirmed by Sanger sequencing (Figures 4d,e and S11). We observed that the level of TPA_S5 and M_TEX_S5 were lower in mutants than those in wild-type plants (Figure 4f). The *Tafbl-7a* mutants also showed reduced SWi and TKW (Figure 4g). Together, the results suggested a role of *TraesCS7A03G0149400* in regulating plant growth traits and yield-related traits in wheat.

We further investigated the genomic sequence of *TraesCS7A03G0149400* in the association population. Surprisingly, two InDels, a 14 bp insertion and an 18 bp deletion, were identified in the first exon after sequencing 15 representative lines in the CC haplotype (Figure S12a). The two InDels in *TraesCS7A03G0149400* cause a frameshift that creates a stop codon (Figure S12b). We genotyped all the 288 lines with markers developed based on the two InDels and successfully obtained results from 242 lines with a total of 22 lines containing the two InDels and all of them are CC genotype at SNP-47107918 (Data S14). Thus, genetic variations in gene coding region of *TraesCS7A03G0149400* influence plant growth phenotype and yield traits. The InDel markers would be used for preferable allele selection in wheat breeding.

Increasing prediction accuracy for yield-related traits using image traits in wheat breeding

Different growth stages are highly correlated, and variation in growth is linked to variation in yield production at maturity (DeWitt *et al.*, 2021; Gowda *et al.*, 2011; Li *et al.*, 2020b; Nadolska-Orczyk *et al.*, 2017; Zhang *et al.*, 2017). However, using phenotypic data obtained at the early growth stages to predict the final yield-related trait remains a major challenge in crop breeding. With recent advances in the non-invasive imaging phenotyping, we are able to monitor the growth dynamics of plants from sowing to harvest. We developed models based on i-traits from different growth stages to predict yield-related traits with two different methods: LASSO (Figures 5 and S13) and random forest models (Figure S14). Both models showed superior performance on PH (plant height), StW (straw weight) and BM (biomass) when only i-traits measured at S6 were included (Figures 5a and S14a), but LASSO model outperformed than RF model on yield traits such as SL, SpN, SN, TKW, etc. Therefore, LASSO method was selected for the following analysis.

The performance of prediction was greatly improved after including i-traits from earlier growth stages (S1–S5). For example, the R value for TN (tiller number) model was 0.4 when using i-traits from S6 stage only, but it was increased to 0.71 after including S5 and S6 stages. The value was further elevated to 0.9 when using i-traits from S4, S5 and S6 stages (Figure 5b). The prediction accuracy for yield–related traits was also improved when including i-traits from earlier growth stages. Correlation coefficient between the experimental data and expected values were generated using the best models (Figures 5c and S15; Data S15). Our results showed that R values for TN and PH are 0.90 and 0.91, respectively. Yield traits, such as SN and GY, showed a relatively lower prediction accuracy with R value at 0.70 and 0.51, respectively (Figure 5b).

We further identified the i-traits contributing to each of the 17 agronomic traits based on the best prediction models (Data S15; Figure 5d). It is not surprising that more i-traits collected at later growth stages (S4, S5 and S6) demonstrated a large contribution of these traits to yield-related traits such as BM, PH, KW (kernel weight), etc (Figure 5d). On the other hand, i-traits from tillering (S2 and S3) stages also contributed to yield-related traits such as SN, GY, KNP, SW, etc (Figure 5d). Considering that plants were

tillering at S2 and S3 stages during winter, the results suggested an influence of winter growth on final yield production in wheat. Together, our results indicated that i-traits collected during the early growth stages largely improved prediction accuracy for final yield, which allows high-throughput selection of targeted plants at early stages thus speed up breeding process.

Discussion

Understanding the complexity of the genetic variants underlying diversity in growth-related traits (e.g. height, width and shape) to maximize yield production is essential in crop breeding. However, growth traits changed dynamically as development progressed and traditional phenotyping throughout plant development is still a bottleneck of genetic studies. In this study, we obtained 190 image-based traits (186 i-traits and four growth parameters) depicting plant growth-related traits and 17 traditional traits (e.g. height, biomass and yield) in a diverse wheat population. The observation and measurement of growth traits and yield traits in one study allows us to dissect the underlying genetic mechanisms of complex traits and investigate the correlation between crop yield potential and vegetative growth.

Most of the previous GWAS studies in wheat were performed with a few traditional traits either using small size of populations or limited numbers of markers (Cao *et al.*, 2020; Jamil *et al.*, 2019; Knoch *et al.*, 2020; Pang *et al.*, 2020). As a result, a few hundreds of marker-traits associations and QTLs spanning on a large genome regions were identified. Compared to previous studies using less than 0.4 million markers from genotyping by sequencing (Pang *et al.*, 2020) or tens of thousands from SNP chips (Würschum *et al.*, 2014), this study generated a total of 12.64 million high-quality SNP/InDel markers using whole genome re-sequencing of 288 wheat lines. The ultra-high-density markers and population with high genetic diversity ensure GWAS analysis in a high-resolution to effectively identify a larger number of QTLs or to directly locate the causal gene. A total of 785 SNP/InDels located in the coding region of candidate genes will facilitate the identification of the causal gene (Data S11), of which the leading SNP (SNP-47107918) causing a missense mutation was located exactly in the second exon of the candidate gene *TraesCS7A03G0149400*. The SNP was significantly associated with 13 i-traits such as TPA and GPA values, which are indicators for plant size. *TraesCS7A03G0149400* encodes a cyclin-like F-box protein which could serve as a protein–protein interaction motif to recruit the substrate-recognition components of the E3 ligase complex (Noir *et al.*, 2015; Xu *et al.*, 2009). Studies have shown that cyclin-like F-box protein plays a role in growth and development of various plant species (Xu *et al.*, 2009). Loss of the F-box protein FBL17 in *Arabidopsis* resulted in reduced plant growth by inhibiting cell proliferation (Noir *et al.*, 2015). Two independent EMS mutants confirmed its role in regulating plant development and yield production in wheat. Furthermore, we identified two novel InDels by sequencing the genomic region of *TraesCS7A03G0149400* in the association population, which could be used as a selection marker for wheat breeding.

Traditional studies examine a few traits once at particular stage; however, most traits are changing dynamically along developmental process. The dynamic regulation of QTLs during development remains mysterious. Using non-destructive high-throughput phenotyping platform (Yang *et al.*, 2014), we collected a number of novel plant growth-related traits (i-traits)
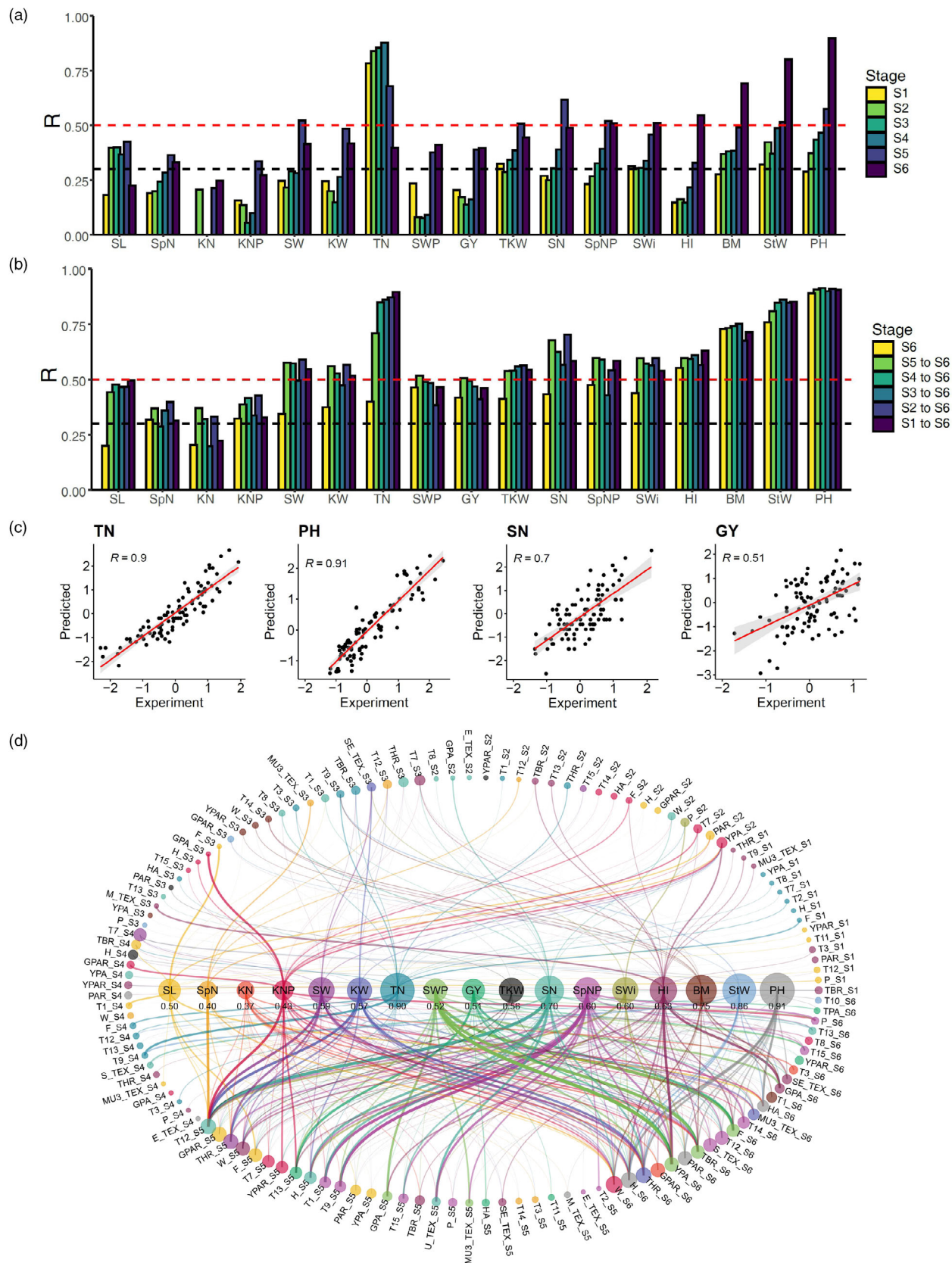
Figure 5 Prediction of yield-related traits using i-traits captured on six growth stages. (a) Prediction models were generated using i-traits collected at single growth stages. (b) Prediction models were generated using i-traits collected at different growth stages. S6, i-traits at S6 stage only; S5-S6, i-traits at S5 and S6 stages; S4-S6, i-traits at S4, S5 and S6 stages; S3-S6, i-traits at S3, S4, S5 and S6 stages; S2-S6, i-traits at S2, S3, S4, S5 and S6 stages; S1-S6, i-traits at S1, S2, S3, S4, S5 and S6 stages. The prediction accuracy was presented as Pearson correlation coefficient value R between predicted value and experiment value. (c) Best prediction models for four yield-related traits were shown. (d) Linkage between i-traits and agronomic traits. The best prediction model for each agronomic trait was derived with i-traits separately. A link between each agronomic trait (inner aligned) and its associated i-traits (outer circle) was created. The thickness of the links is determined by the absolute values of the coefficients of each independent variable (i-trait) in the linear regression prediction model. Each agronomic trait was assigned a colour, and the i-traits with the highest absolute values of coefficients was assigned the same colour. The dot size for each agronomic trait is determined by the number of i-traits in its prediction model. The dot size for each i-trait represents the number of its appearance in prediction models.

at multiple stages that were not measurable using traditional methods, e.g. plant compactness and green leaf area. The growth habit of wheat is largely controlled vernalization genes such as *VERNALIZATION 1* (*VRN1*), *VRN2* and *VRN3* (Distelfeld *et al.*, 2009; Trevaskis *et al.*, 2007; Yan *et al.*, 2004). *Vrn1-5A* has been shown to be associated with tiller number and prostrate plant type in wheat (Iwaki *et al.*, 2000; Marone *et al.*, 2020); thus, changes in tiller number and prostrate will probably influence the plant type. In this study, *Vrn1-5A* was identified as a candidate gene for TBR_S3, an i-trait depicting plant compactness at tillering stage (S3). *Vrn1-5D* (also called *Vrn-D1*), which promotes heading and flowering dates in wheat (Zhang *et al.*, 2012, 2018), is located in the pleiotropic *QTL168* associated with 50 i-traits at S2, S3, S4 or S5 stages. *VRN3* is a close homologue to the *FLOWERING LOCUS T* in *Arabidopsis* which promotes flowering under long days (Yan *et al.*, 2006). In consistent, we identified *Vrn3-7A* as a candidate gene in two QTLs, qPAR_S6_7A and qTBR_S6_7A, which were associated with PAR and TBR i-traits at grain filling stage (S6) when days are longer in spring. These results suggested a major role of vernalization genes in mediating plant growth phenotype in a stage-specific manner. With the fact that growth traits change dynamically as development progressed and are governed by numerous loci, our study demonstrated the dynamics of QTLs during plant development at multiple growth stages via high-throughput phenotyping and genotyping in wheat.

Accurate prediction of yield potential is of great significance in crop breeding. Previous studies have shown that yield, as a complex trait at harvest, is affected by early plant growth and biomass accumulation (DeWitt *et al.*, 2021; Gowda *et al.*, 2011; Knoch *et al.*, 2020; Li *et al.*, 2020b; Nadolska-Orczyk *et al.*, 2017). We observed that some QTLs for i-traits collected at growth stages co-localized with QTLs for yield-related traits at harvest. Prediction models were derived from multiple growth stages and greatly increased the prediction accuracy for yield-related traits. Both results indicated that earlier plant performance contributes to final yield production. It has been proposed that genes are selectively expressed at different growth stages and the genetic control of development at different developmental stages determines to the final yield (Fu *et al.*, 2022; Zhu, 1995). How these genetic loci are hierarchically regulated and coordinated at different growth stages are largely unknown. We further dissected a number of i-traits contributing to yield production based on machine learning models. These i-traits were thus considered as developmental modules, and their QTLs were regarded as genetic modules that dictate crop performance and yield. Hence, our results indicated that pleiotropy or dynamic regulation of QTLs at multiple stages interacted with each other may be responsible for the linkage between plant growth and yield production. Importantly, we demonstrated that the yield-related traits are predictable and dissectible using i-traits collected during plant development, which provides possibility for selecting desired trait combinations to maximize crop yield or predicting yield potential in breeding programs.

Experimental procedures

Plant materials and growth condition

In this study, a wheat association panel with 288 inbred lines (Data S1) was planted in the high-throughput phenotyping platform (HTP) (Yang *et al.*, 2014) with three biological replicates for each line. Seeds were sown directly in pots with 4.5 kg soil on October 25 2019, and one wheat plant was kept for each pot. Fertilization was carried out at sowing and tillering stage (60 kg of water +370.68 g of carbamide +330.76 g of potassium dihydrogen phosphate +94.24 g of potassium chloride, to be fully dissolved with 150 mL of liquid fertilizer for each plant per time) (Zhang *et al.*, 2017). All the plants were screened at six time points (S1-S6 represent 44, 57, 71, 84, 120 and 180 days after sowing), which corresponding to seedling (S1), early tillering (S2), late tillering (S3), stem elongation (S4), heading (S5) and grain filling stage (S6). The growth conditions and trait collection were shown in Figure S3.

In order to confirm the involvement of the candidate genes, we obtained EMS mutants in the AK58 background and backcrossed with AK58 twice to purify the genetic background. The putative target mutation sites were verified by Sanger sequencing using gene-specific primers (Data S16). Homozygous mutants were used for i-trait and agronomic trait detection. Individual mutants were grown in HTP with at least three replicates.

High-throughput phenotyping and statistical analysis

RGB images of the wheat plants were captured with high-throughput phenotyping facility (Yang *et al.*, 2014). Six side view images were taken at each time when the pot was automatically rotated by 60°. After image processing, 31 image-based digital traits (i-traits) were extracted (Data S3 and S5). Morphological traits were extracted based on the height, width, area and colour of the plant which depict plant height, size and shape. Texture traits were derived from the brightness, grey level, gradient of pixels, which represent the grey level distribution and the spatial relationship of pixels in the image (Yang *et al.*, 2014).

To determine the best model for TPA (total projected area) and H (height of the bounding rectangle), four growth models (including linear, exponential, quadratic and power models) were evaluated using the coefficient of determination (R^2) and the mean absolute percentage error (MAPE). The exponential model was selected to fit TPA and H at six time points. The parameters generated from exponential model including TPA_a, TPA_b, H_a and H_b were used as i-traits for GWAS.

At the end of the experiment, 17 yield-related traits were measured manually, including plant height (PH), tiller number (TN) and spike number (SN) (Data S3 and S5). After harvest, the spikes were collected for spike-related traits measurement. For each spike, spike weight (SW), spike length (SL), kernel number per spike (KN) and kernel weight (KW) (Liu *et al.*, 2018) were collected. Spike weight per plant (SWP), kernel number per plant (KNP) and grain yield (GY, measured as kernel weight per plant) were calculated by sum the corresponding spikes traits from all spikes from the plant. Thousand kernel weight (TKW) was generated by total grains from one plant, generally including more than 200 seeds. Straw weight was measured after drying in a convection oven at 45 °C for 48 h.

Heritability was calculated for each trait as follows:

$$H^2 = \sigma_G^2 / [\sigma_G^2 + \sigma_e^2/r]$$

where σ_G^2 is the genotypic variance, σ_e^2 is the error variance, and r is the number of replicates. The σ_G^2 and σ_e^2 were estimated in R environment.

Trait correlation analysis and heatmap were conducted with *Hmisc* and *pheatmap* packages in R environment. For growth model selection, linear model was tested with *lm* function and the rest three models were performed with *nlr* package.

Genotyping and population structure analysis

Genomic DNA of 288 core wheat lines was extracted and sequenced on MGI 2000 using the standard protocol. Raw reads were processed for quality control and trimming with FastQC tool (v0.11.7; Babraham Bioinformatics). The clean resulted high-quality reads were aligned to wheat reference genome, IWGSC_ref 2.1 (Zhu *et al.*, 2021) with Burrows-Wheeler Aligner (BWA) software (Li and Durbin, 2009). Then, the BAM alignment files were subsequently generated in *samtools* (Li *et al.*, 2009). The resulting variant call format (VCF) file was filtered with read depth of 1269, miss ratio < 20% and minor allele frequency (MAF) > 0.05; the SNP annotation was performed according to the reference genome, IWGSC_ref 2.1 with SnpEff (Cingolani *et al.*, 2012). PopLDdecay (Zhang *et al.*, 2019) was employed for linkage disequilibrium (LD) decay evaluation. The phylogenetic analysis and kinship were conducted with Plink (Purcell *et al.*, 2007) and KING (Manichaikul *et al.*, 2010), respectively. The population structure was estimated with Admixture (Alexander *et al.*, 2009).

Genome-wide association analysis candidate gene selection

After quality control, 12 182 649 single nucleotide polymorphisms (SNPs) and 456 364 insertion and deletions (InDel) were obtained and subsequently used for genome-wide association analysis (GWAS) with 207 traits including 186 i-traits, 4 growth parameters and 17 agronomic traits. GWAS was performed using FaST-LMM program (Lippert *et al.*, 2011) with a mixed model. The population structure vector (K = 9) and a kinship matrix were used to correct population structure and family relatedness. A total of 12 639 013 markers were involved in GWAS. To reduce false negatives due to Bonferroni correction (1/number of markers), significant markers trait association were selected with *P*-value < 1E-7 for all traits in this study. The significantly associated markers were extended to LD-based QTLs using $r^2 = 0.5$ as threshold, as previously described (Pang *et al.*, 2020). For each trait, all significant SNPs within

overlapping QTL intervals were categorized as an associated locus and represented by lead markers (that with the lowest *P* value) and QTL close to each other (distance < 2 Mb) were merged. QTL with spanned region larger than 50 Mb were excluded. The variance explained by the locus was estimated with *lm* function in R software.

The annotated genes within the identified QTL were extracted from the IWGSC_ref 2.1 as candidate genes. In addition, genes adjacent lead markers within 3 Mb both upstream and downstream are also included as potential candidate genes. These adjacent genes were selected by: (1) non-synonymous markers were detected on the genes; (2) these non-synonymous markers were significantly associated with the trait at a lower threshold, *P*-value < 1E-6. The corresponding rice orthologous genes were identified by *blastp* function from BLAST+ suite (Camacho *et al.*, 2009) with Evalue < 1E-5 as threshold. The rice protein sequences were downloaded from MSU Rice Genome Annotation Project Release 7 (Kawahara *et al.*, 2013). The keywords for the founded rice genes were annotated with funRiceGenes (Yao *et al.*, 2018).

Machine learning methods for the prediction of agronomic traits

Two machine learning regression models, least absolute shrinkage and selection operator (LASSO) and Random Forest (RF), were assessed for yield-related traits prediction based on i-traits collected at six stages. In LASSO method, generalized linear models were developed based on i-traits in several growth stages:

$$y = a + B \times X.$$

where y is the target agronomic trait; a is the intercept; B is a vector of slope; X is the i-traits vector with a length of 31 × n (number of growth stages). The models were estimated with *glnmet* package based on a least absolute shrinkage and selection operator (LASSO) method (Friedman *et al.*, 2010). In RF method, the *caret::train ()* function was used and models were developed based on i-traits collected at different stages. The phenotype data was divided into threefolds (K = 3), with twofolds for model training and onefold for testing. The performance of prediction models was evaluated with cross-validation. Pearson's correlation (R) was calculated between predicted values and experimental values for testing set.

Genotyping of *Vrn1* genes and *TraesCS7A03G0149400* with PCR based primers

Vrn1-5A, *Vrn1-5B* and *Vrn1-5D* were genotyped using gene-specific primers as previously described (Yan *et al.*, 2004). Genomic DNA from 288 wheat accessions was used as template and PCRs were performed using Taq DNA Polymerase (Vazyme, Nanjing, China) with primers listed in Data S16. The PCR products were analysed with electrophoresis through agarose gels for *Vrn1* genes. Gene-specific primers were designed for the detection of InDels in *TraesCS7A03G0149400* (Data S16). The PCR products were analysed with electrophoresis through polyacrylamide gels for the association population.

Accession numbers

All the genome sequencing data generated in this study have been deposited in the public database of the China National Genebank (https://db.cngb.org/cnsa) under the accession number CNP0003712.

Competing interests

The authors declare no competing interests.

Author contributions

Q.L. and J.G. conceived and designed the study. J.G. and X.H. performed the experiments and data analysis. H.Y., H.L., C.G., G.C., H.F., Z.J., Z.G., J.F., J.Z., Y.C., B.Y. and Z.P. helped collecting data or analysed the data. WN.Y, WH.Y. and L.G. supported the study, provided materials, involved in manuscript writing and discussion. J.G. and Q.L. analysed the data and wrote the manuscript. D.X., J.J., H.S., H.M., C.L. and W.C. provided constructive suggestions and revised the manuscript. All authors read and approved the final manuscript.

References

Alexander, D.H., Novembre, J. and Lange, K. (2009) Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* **19**, 1655–1664.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K. and Madden, T.L. (2009) BLAST+: Architecture and applications. *BMC Bioinformatics* **10**, 421.

Camargo, A.V., Mackay, I., Mott, R., Han, J., Doonan, J.H., Askew, K., Corke, F. et al. (2018) Functional mapping of quantitative trait loci (QTLS) associated with plant performance in a wheat MAGIC mapping population. *Front. Plant Sci.* **9**, 887.

Cao, S., Xu, D., Hanif, M., Xia, X. and He, Z. (2020) Genetic architecture underpinning yield component traits in wheat. *Theor. Appl. Genet.* **133**, 1811–1823.

Cingolani, P., Platts, A., Wang, L.L., Coon, M., Nguyen, T., Wang, L., Land, S.J. et al. (2012) A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. *Fly (Austin).* **6**, 80–92.

Cui, F., Fan, X., Chen, M., Zhang, N., Zhao, C., Zhang, W., Han, J. et al. (2016) QTL detection for wheat kernel size and quality and the responses of these traits to low nitrogen stress. *Theor. Appl. Genet.* **129**, 469–484.

DeWitt, N., Guedira, M., Lauer, E., Murphy, J.P., Marshall, D., Mergoum, M., Johnson, J. et al. (2021) Characterizing the oligogenic architecture of plant growth phenotypes informs genomic selection approaches in a common wheat population. *BMC Genomics* **22**, 402.

Distelfeld, A., Li, C. and Dubcovsky, J. (2009) Regulation of flowering in temperate cereals. *Curr. Opin. Plant Biol.* **12**, 178–184.

Dobrovolskaya, O., Pont, C., Sibout, R., Martinek, P., Badaeva, E., Murat, F., Chosson, A. et al. (2015) Frizzy panicle drives supernumerary spikelets in bread wheat. *Plant Physiol.* **167**, 189–199.

Fan, X., Cui, F., Ji, J., Zhang, W., Zhao, X., Liu, J.J., Meng, D. et al. (2019) Dissection of pleiotropic QTL regions controlling wheat spike characteristics under different nitrogen treatments using traditional and conditional QTL mapping. *Front. Plant Sci.* **10**, 187.

FAO (2020) *Crop Prospects and Food Situation #1, March 2020*. Rome: FAO. https://doi.org/10.4060/ca8032en

Friedman, J., Hastie, T. and Tibshirani, R. (2010) Regularization paths for generalized linear models via coordinate descent. *J. Stat. Softw.* **33**, 1–22.

Fu, Y., Zhao, H., Huang, J., Zhu, H., Luan, X., Bu, S., Liu, Z. et al. (2022) Dynamic analysis of QTLs on plant height with single segment substitution lines in rice. *Sci. Rep.* **12**, 5465.

Gao, F., Ma, D., Yin, G., Rasheed, A., Dong, Y., Xiao, Y., Xia, X. et al. (2017) Genetic progress in grain yield and physiological traits in Chinese wheat cultivars of southern Yellow and Huai Valley since 1950. *Crop. Sci.* **57**, 760–773.

Gowda, M., Hahn, V., Reif, J.C., Longin, C.F.H., Alheit, K. and Maurer, H.P. (2011) Potential for simultaneous improvement of grain and biomass yield in Central European winter triticale germplasm. *F. Crop. Res.* **121**, 153–157.

Hong, M.J., Kim, J.B., Seo, Y.W. and Kim, D.Y. (2020) F-box genes in the wheat genome and expression profiling in wheat at different developmental stages. *Genes (Basel)* **11**, 1–16.

Iwaki, K., Nakagawa, K., Kuno, H. and Kato, K. (2000) Ecogeographical differentiation in east Asian wheat, revealed from the geographical variation of growth habit and Vrn genotype. *Euphytica* **111**, 137–143.

Jamil, M., Ali, A., Gul, A., Ghafoor, A., Napar, A.A., Ibrahim, A.M.H., Naveed, N.H. et al. (2019) Genome-wide association studies of seven agronomic traits under two sowing conditions in bread wheat. *BMC Plant Biol.* **19**, 149.

Kawahara, Y., de la Bastide, M., Hamilton, J.P., Kanamori, H., Mccombie, W.R., Ouyang, S. et al. (2013) Improvement of the oryza sativa nipponbare reference genome using next generation sequence and optical map data. *Rice* **6**, 3–10.

Knoch, D., Abbadi, A., Grandke, F., Meyer, R.C., Samans, B., Werner, C.R., Snowdon, R.J. et al. (2020) Strong temporal dynamics of QTL action on plant growth progression revealed through high-throughput phenotyping in canola. *Plant Biotechnol. J.* **18**, 68–82.

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, B., Chen, L., Sun, W., Wu, D., Wang, M., Yu, Y., Chen, G. et al. (2020a) Phenomics-based GWAS analysis reveals the genetic architecture for drought resistance in cotton. *Plant Biotechnol. J.* **18**, 2533–2544.

Li, H., Feng, H., Guo, C., Yang, S., Huang, W., Xiong, X., Liu, J. et al. (2020b) High-throughput phenotyping accelerates the dissection of the dynamic genetic architecture of plant growth and yield improvement in rapeseed. *Plant Biotechnol. J.* **18**, 2345–2353.

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.

Liu, J., Xu, Z., Fan, X., Zhou, Q., Cao, J., Wang, F. et al. (2018) A genome-wide association study of wheat spike related traits in China. *Front. Plant Sci.* **871**, 1584.

Manichaikul, A., Mychaleckyj, J.C., Rich, S.S., Daly, K., Sale, M. and Chen, W.M. (2010) Robust relationship inference in genome-wide association studies. *Bioinformatics* **26**, 2867–2873.

Marone, D., Rodriguez, M., Saia, S., Papa, R., Rau, D., Pecorella, I. et al. (2020) Genome-wide association mapping of prostrate/erect growth habit in winter durum wheat. *Int. J. Mol. Sci.* **21**, 394.

Muraya, M.M., Chu, J., Zhao, Y., Junker, A., Klukas, C., Reif, J.C. and Altmann, T. (2017) Genetic variation of growth dynamics in maize (*Zea mays* L.) revealed through automated non-invasive phenotyping. *Plant J.* **89**, 366–380.

Nadolska-Orczyk, A., Rajchel, I.K., Orczyk, W. and Gasparis, S. (2017) Major genes determining yield-related traits in wheat and barley. *Theor. Appl. Genet.* **130**, 1081–1098.

Noir, S., Marrocco, K., Masoud, K., Thomann, A., Gusti, A., Bitrian, M., Schnittger, A. et al. (2015) The control of Arabidopsis thaliana growth by cell proliferation and endoreplication requires the F-box protein FBL17. *Plant Cell* **27**, 1461–1476.

O'Brien, J.A., Vega, A., Bouguyon, E., Krouk, G., Gojon, A., Coruzzi, G. and Gutiérrez, R.A. (2016) Nitrate Transport, Sensing, and Responses in Plants. *Mol. Plant* **9**, 837–856.

Ohkubo, Y., Kuwata, K. and Matsubayashi, Y. (2021) A type 2C protein phosphatase activates high-affinity nitrate uptake by dephosphorylating NRT2.1. *Nat. plants* **7**, 310–316.

Pang, Y., Liu, C., Wang, D., St. Amand, P., Bernardo, A., Li, W. *et al.* (2020) High-resolution genome-wide association study identifies genomic regions and candidate genes for important agronomic traits in wheat. *Mol. Plant* **13**, 1311–1327.

Pearce, S., Saville, R., Vaughan, S.P., Chandler, P.M., Wilhelm, E.P., Sparks, C.A., al-Kaff, N. *et al.* (2011) Molecular characterization of Rht-1 dwarfing genes in hexaploid wheat. *Plant Physiol.* **157**, 1820–1831.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., 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.

Shi, X., Cui, F., Han, X., He, Y., Zhao, L., Zhang, N., Zhang, H. *et al.* (2022) Comparative genomic and transcriptomic analyses uncover the molecular basis of high nitrogen-use efficiency in the wheat cultivar Kenong 9204. *Mol. Plant* **15**, 1440–1456.

Tester, M. and Langridge, P. (2010) Breeding technologies to increase crop production in a changing world. *Science* **327**, 818–822.

Trevaskis, B., Hemming, M.N., Dennis, E.S. and Peacock, W.J. (2007) The molecular basis of vernalization-induced flowering in cereals. *Trends Plant Sci.* **12**, 352–357.

Watt, M., Fiorani, F., Usadel, B., Rascher, U., Muller, O. and Schurr, U. (2020) Phenotyping: New Windows into the Plant for Breeders. *Annu. Rev. Plant Biol.* **71**, 689–712.

Wu, X., Feng, H., Wu, D., Yan, S., Zhang, P., Wang, W., Zhang, J. *et al.* (2021) Using high-throughput multiple optical phenotyping to decipher the genetic architecture of maize drought tolerance. *Genome Biol.* **22**, 185.

Würschum, T., Liu, W., Busemeyer, L., Tucker, M.R., Reif, J.C., Weissmann, E.A., Hahn, V. *et al.* (2014) Mapping dynamic QTL for plant height in triticale. *BMC Genet.* **15**, 59.

Xiong, H., Zhou, C., Fu, M., Guo, H., Xie, Y., Zhao, L., Gu, J. *et al.* (2022) Cloning and functional characterization of Rht8, a "Green Revolution" replacement gene in wheat. *Mol. Plant* **15**, 373–376.

Xu, G., Ma, H., Nei, M. and Kong, H. (2009) Evolution of F-box genes in plants: Different modes of sequence divergence and their relationships with functional diversification. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 835–840.

Yan, L., Loukoianov, A., Tranquilli, G., Helguera, M., Fahima, T. and Dubcovsky, J. (2003) Positional cloning of the wheat vernalization gene VRN1. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 6263–6268.

Yan, L., Helguera, M., Kato, K., Fukuyama, S., Sherman, J. and Dubcovsky, J. (2004) Allelic variation at the VRN-1 promoter region in polyploid wheat. *Theor. Appl. Genet.* **109**, 1677–1686.

Yan, L., Fu, D., Li, C., Blechl, A., Tranquilli, G., Bonafede, M., Sanchez, A. *et al.* (2006) The wheat and barley vernalization gene VRN3 is an orthologue of FT. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 19581–19586.

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.

Yao, W., Li, G., Yu, Y. and Ouyang, Y. (2018) funRiceGenes dataset for comprehensive understanding and application of rice functional genes. *Gigascience* **7**, 1–9.

Zhang, J., Wang, Y., Wu, S., Yang, J., Liu, H. and Zhou, Y. (2012) A single nucleotide polymorphism at the Vrn-D1 promoter region in common wheat is associated with vernalization response. *Theor. Appl. Genet.* **125**, 1697–1704.

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, X., Chen, J., Yan, Y., Yan, X., Shi, C., Zhao, L. and Chen, F. (2018) Genome-wide association study of heading and flowering dates and construction of its prediction equation in Chinese common wheat. *Theor. Appl. Genet.* **131**, 2271–2285.

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.

Zhang, X., Jia, H., Li, T., Wu, J., Nagarajan, R., Lei, L., Powers, C. *et al.* (2022) TaCol-B5 modifies spike architecture and enhances grain yield in wheat. *Science* **376**, 180–183.

Zhu, J. (1995) Analysis of conditional genetic effects and variance components in developmental genetics. *Genetics* **141**, 1633–1639.

Zhu, T., Wang, L., Rimbert, H., Rodriguez, J.C., Deal, K.R., De Oliveira, R. *et al.* (2021) Optical maps refine the bread wheat Triticum aestivum cv Chinese Spring genome assembly. *Plant J.* **107**, 303–314.

Supporting information

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

Figure S1. Principal component biplots of association population.

Figure S2. Analysis of the population structure.

Figure S3. Growth conditions and phenotyping for the wheat association panel.

Figure S4. Correlation network between i-traits and agronomic traits.

Figure S5. Summary of QTLs.

Figure S6. Pleiotropic QTLs detected in this study.

Figure S7. Pleiotropic QTLs associated with both agronomic traits and i-traits.

Figure S8. Analysis of *Vrn1-5A*, *Vrn1-5B* and *Vrn1-5D* alleles in the wheat association panel.

Figure S9. Phylogenetic tree for F-box proteins in wheat.

Figure S10. Haplotype analysis of SNP-47107918 in the wheat association panel.

Figure S11. Confirmation of SNP-47107918 and point mutations in *Tafbl-7a* mutants.

Figure S12. Sequence analysis of two InDels and SNP-47107918 in *TaFBL-7A*.

Figure S13. Prediction of yield-related traits using i-traits with LASSO models.

Figure S14. Prediction of yield-related traits using i-traits with Random Forest models.

Figure S15. Pearson correlations between predicted values and experimental values for agronomic traits.

Data S1. The 288 wheat accessions used in this study.

Data S2. Genetic diversity of markers identified by genome re-sequencing.

Data S3. Description of agronomic traits and image traits.

Data S4. Growth models for height of the bounding rectangle (H) and total projected area (TPA).

Data S5. Phenotypic data collected in this study.

Data S6. Phenotypic variation for agronomic and image traits.

Data S7. Correlations between agronomic traits and image traits.

Data S8. QTLs identified in this study.

Data S9. Comparisons to previously reported QTLs in wheat.

Data S10. Candidate genes identified in QTLs.

Data S11. Non-synonymous markers in QTLs.

Data S12. Pleiotropic QTLs identified in this study.

Data S13. Genotyping of *Vrn1* genes in 288 wheat accessions.

Data S14. Genotyping *TraesCS7A03G0149400* in wheat association population.

Data S15. Best prediction models for agronomic traits using image traits.

Data S16. Primers used in this study.