Dissecting the phenotypic components and genetic architecture of maize stem vascular bundles using high-throughput phenotypic analysis

Ying Zhang1,†, Jinglu Wang1,†, Jianjun Du1,†, Yanxin Zhao2 (✉), Xianju Lu1, Weiliang Wen1, Shenghao Gu1, Jiangchuan Fan1, Chuanyu Wang1, Sheng Wu1, Yongjian Wang1, Shengjin Liao1, Chunjiang Zhao1,* and Xinyu Guo1,*

1 Beijing Key Lab of Digital Plant, National Engineering Research Center for Information Technology in Agriculture, Beijing Research Center for Information Technology in Agriculture, Beijing Academy of Agriculture and Forestry Sciences, Beijing, China
2 Beijing Key Laboratory of Maize DNA Fingerprinting and Molecular Breeding, Maize Research Center, Beijing Academy of Agriculture and Forestry Sciences, Beijing, China

Keywords: maize stem, vascular bundle, micro-CT, micro-phenotype, genome-wide association studies, MaizeSPD.

Introduction

Maize (Zea mays ssp. mays) was domesticated from its wild ancestor, teosinte (Zea mays ssp. parviglumis) nearly 6000 to 10 000 years ago in southwestern Mexico (Doebly, 1990, 2004). The widely cultivation and strong adaptation allow maize to be the largest productive crop in the worldwide, which is important in the satisfying global food demand and safeguarding world food security (Liu et al., 2015; USDA FAS, 2013). The yield of maize closely relies on the source–sink relationship, and the ideal state of that is plenty in source, rich in sink and efficient in flow. In recent year, with the genetic improvement of source–sink traits and great progress in cultivation measures, more and more attentions have been paid in the role of ‘flow’ in yield formation (Wang et al., 2011). Vascular bundle, the important ‘flow’ which links source and sink, is responsible for both the delivery of water, essential mineral nutrients, sugars and amino acids among organs and decides the transportation efficiency of photosynthetic products, water and essential mineral nutrients (Housley and Peterson, 1982). As an effective long-distance transport system of source—translocation–sink, correlations between vascular bundle performance and grain yield have been reported in crops (Chen et al., 2004; Cui et al., 2003; Nátrová, 1991; Peterson et al., 1982). Studies have shown the number, size and role of vascular bundle directly affected the transporting efficiency of assimilates from the source to kernels and eventually as important limiting causes effecting crop yield (Housley and Peterson, 1982; Huang et al., 2016; Nátrová, 1985; Zhai et al., 2018). Vascular bundle traits are so important that it is critical to make unremitting efforts to explore genetic mechanism. In Arabidopsis, many genes took part in vascular bundle patterning have been identified, such as PHB, PHV, ATHB15 and REV (Du and Wang, 2015; McConnell et al., 2001; Zhong and Ye, 2004). In rice, many quantitative trait loci (QTLs) for vascular bundle have been identified (Bai et al., 2012; Fei et al., 2019; Zhai et al., 2018), and candidate genes such as APO1, ABV, DEP1 and NAL1 have been reported (Fei et al., 2019; Fujita et al., 2013; Qi et al., 2008; Terao et al., 2010). In maize, Huang et al. (2016) identified quantitative trait loci (QTL) for the number of vascular bundles in the uppermost internode of maize
In recent years, with the rapid development of high-density single nucleotide polymorphism (SNP) genotyping and the next-generation sequencing (NGS) technologies, genome-wide association study (GWAS) has become a powerful tool to dissect the genetic basis for the quantitative variation of complex traits in crops (Chen et al., 2019; Xiao et al., 2017). For maize, since the release of the B73 reference genome (Schnable et al., 2009), many agronomic important traits, such as plant height (Dell’Acqua et al., 2015; Farfan et al., 2015; Li et al., 2016a,b; Peiffer et al., 2014; Riedelsheimer et al., 2012; Wang et al., 2019; Weng et al., 2011; Yang et al., 2014a,b), flowering time (Buckler et al., 2009; Farfan et al., 2015; Hung et al., 2012; Li et al., 2016a,b; Van Inghelandt et al., 2012; Yang et al., 2013, 2014a,b), ear height (Dell’Acqua et al., 2015; Farfan et al., 2015; Li et al., 2016a,b; Peiffer et al., 2014; Yang et al., 2014a,b) and grain size (Dell’Acqua et al., 2015; Yang et al., 2014a,b), have been dissected through GWASs. GWAS application in agricultural traits of maize provides useful reference for revealing the phenotypic traits diversity and genetic architecture of vascular bundles in maize stem. However, because of the need for accurate identification of microscopic phenotypes for large amounts of maize and a lack of high-throughput and effective microphenotyping detection methods, few GWASs for ‘flow’ traits have been performed in large population of maize inbred lines.

With the rapid development of functional genomics and molecular breeding, the ability to quickly screen thousands of lines for targeted phenotypic traits is becoming more and more important (Fiorani and Schurr, 2013). Manually counting vascular bundle number was a strenuous and tedious work, and errors in the measurement were unavoidable. What is more, many anatomical traits of vascular bundles could not be detected by manual test, which seriously affected develops in maize genomics for these important characters of vascular bundles. To bridge this gap, progress in the development of high-throughput phenotyping technology is required (Yang et al., 2013a,b). In recent years, several tools for automated detecting phenotypes of stem have become available (Du et al., 2016; Heckwolf et al., 2015; Legland et al., 2014; Zhang et al., 2013). However, much more robust and accurate identification methods are urgently needed that are suitable for large populations of maize.

In this study, given the rich genetic variation in maize natural population (Yang et al., 2011), we developed a standard process for stem micro-CT data acquisition and automatic CT image process pipeline to extract micro-phenotypic traits. The stems of maize natural population panel containing 480 inbred lines were phenotyped at the silking stage, and 48 traits were automatically extracted by CT image processing pipeline at one time. Based on representative phenotypic data, the phenotypic properties of stem vascular bundles of maize diverse natural population were analysed. And we established the first database for stem micro-phenotypes, MaizeSPD, which stored 554 pieces of basic information for maize inbred lines, 523 pieces of experimental information, 1008 pieces of CT scanning images and processed images, and 24 192 pieces of micro-phenotypic data. Finally, GWASs were conducted to reveal the natural genetic variation and to dissect the genetic architecture of vascular bundles; a total of 1562 significant SNPs were identified for 30 stem micro-phenotypic traits, and 84 unique genes of 20 traits such as VBNum, VBAvArea, IZVBNum and PZVBDensity were detected. The results presented here will advance our knowledge about phenotypic trait components of stem vascular bundles and will provide useful information for understanding the genetic controls of vascular bundle formation and development.

**Results**

**Extraction of vascular bundle phenotypes**

How to quantify the traits of maize stem and vascular bundles is a challenge, also an opportunity of standardize the protocol of data gaining and image processing from micro-phenotyping of maize stem. Based on the morphological characteristics of stem vascular bundles and previous research (Du et al., 2016), we developed automated image analysis pipeline that suitable for large-population CT images to extract the micro-phenotypic traits of maize stems automatically. All image processing and analysis steps were conducted in Visual studio C++ and OpenCV. The function modules of VesselParser 4.0 software (NERCITA, Beijing, China) included data management module, method parameter module, phenotyping calculation module and statistic analysis module (Figure 1 and Figure S1), and the flow chart outlines of the image analysis pipeline are summarized in Figure 1.

During the maize growth and development, the substance accumulation of maize stem shows as the gradually increasing average HU values (i.e. intensity values) and changing distribution (i.e. connectivity relationship). According to the observable intensity and distribution differences between stem substances, the entire stem slice could be reasonably divided into three functional zones, that is epidermis zone, periphery zone and inner zone, corresponding to the anatomy of stem namely epidermis, periderm and pith. In our knowledge, the zonation of stem tissue is difficult to be accurately measured by manual work owing to the boundary ambiguity, and there are no more automated segmentation methods. Once the functional zones were segmented, vascular bundles in each zone were extracted and phenotypic traits of that were calculated. Using VesselParser 4.0 software, 48-item phenotypic traits were automatically extracted and calculated at one time, including 18-nondimensional morphological and 30-dimensional geometrical features, and most of which were difficult to measure and mark through manual work. The list and abbreviations of these 48 traits are shown in Table S1. The average calculation time of one image is about 30 s, and large quantities of images can be conducted batch processing.

**Phenotypic variations of vascular bundles between natural population subgroup**

Based on the phenotypic data gained by VesselParser 4.0 pipeline, we furtheranalysed the variation in vascular bundle traits of the third internode of maize stem between natural population subgroup. Wide phenotypic variations in vascular bundle size, morphology, number, distribution density and other characteristics in cross section and functional zones (epidermal/periphery/inner zones) were observed in NP population (Figure 2), and 48 phenotypic parameters in NP population are listed in Table S2. The average area of circularity of inner zone (IZCir) ranging from 0.01863 to 0.1814 with an average of 0.05945 has the highest maximum change of 9.74-fold, followed by average area of inner zone vascular bundles (IZVBAvArea, 7.30-fold), number of inner...
Maize stem vascular bundle phenotypic traits and genetic architecture analysis

Figure 1 Function modules of VesselParser 4.0 software, including data management module, method parameter module, phenotyping computation module and statistic analysis module. The phenotyping computation module is further demonstrated a flowchart outlines of the image analysis pipeline.

zone vascular bundles (IZVBNumb, 6.33-fold), area of periphery zone (PZArea, 6.25-fold) and area of inner zone (IZArea, 6.20-fold). The rectangularity of stem cross section (CSRRect), rectangularity of epidermis zone (EZRect) and rectangularity of periphery zone (PZRect) had the lowest change of 1.03-fold. Moreover, the frequency distribution of 48-item phenotypic traits in NP population showed a continuous variation (Figure S2), which suggested that vascular bundle phenotypes were the typical quantitative traits controlled by polygenes.

Based on phenotypic data, clustering analysis of 48 phenotypic traits was conducted with hierarchical clustering using the Pearson correlation as a distance metric. The 48 accessions were clustered into four major groups, and the resulting dendrogram is shown in Figures S3 and S4. Geometric and morphological characters of stem cross section and functional zones, and quantitative character of vascular bundles were gathered into group I, including 25 phenotypic parameters. Group II was consisted of seven phenotypic parameters, which were related to vascular bundle area characteristics. Distribution properties of vascular bundle were classified into group II, containing three phenotypic parameters. Group IV was composed of the remaining 12 phenotypic parameters, representing geometric and morphological characters of stem cross section and functional zones. Cluster analysis results above provided the basis for selecting the most representative and biologically intuitive traits from one cluster for GWAS and follow-up analysis.

The heritability ($h^2$) of a trait is one of the key parameters used for making decisions concerning the design and selection of plant breeding schemes (Chen and Lübberstedt, 2010; Holland et al., 2003). Next, heritability was calculated for each trait of four categories. Forty-eight phenotypic parameters showed different heritability patterns, ranging from 0.128 to 0.836 (Figure 3). About 77% (37 items parameters) had heritability >0.5 and more than 60% (30 items) of the parameters had heritability >0.7, indicating that variability in stem micro-trait is governed in a large part by genetic factors. The phenotypic traits from groups 1, 2 and 3 showed high heritability, whereas phenotypic traits from group 4 had low heritability, which might be due to the lower genetic variation of these traits. So far, according to the clustering analysis and heritability values, we selected 30 phenotypic indicators with heritability higher than 0.7, allowing for the natural variation analysis of phenotypic traits in subpopulation and identification of SNPs controlling their expression in revealing plant genetic regulation mechanisms.

Based on the 30 phenotypic indicators with heritability higher than 0.7, an ANOVA was used to discover whether differences exist between the different subpopulations of NP population (TST, NSS, SS and mixed). We found that in addition to the traditional index of vascular bundle number, the phenotypic indexes with the most significant differences ($P \leq 0.001$) between subpopulations also included VBAvArea, IZVBNumb, PZVBAvArea and PZVBDensity (as shown in Figure 4). The number of stem vascular bundles (VBNumb) was much higher for TST than for NSS, SS and mixed, but average area of stem vascular bundles (VBAvArea) was much lower for TST than for NSS, SS and mixed. The phenotypic differences of vascular bundles in inner zone were as the same as that in the stem cross section. For the periphery zone, there was no significant difference in vascular bundle number, but the differences in the average area of vascular bundles and vascular bundle density in that zone were extremely significant. For example, vascular bundle density in periphery zone (PZVBDensity) was much higher for TST than for NSS, SS and mixed. These results reflected the morphological structure and distribution characteristics of vascular bundles among different genotypes; for example, the number of stem vascular bundles tends to be much more and higher vascular bundle density, but the area of vascular bundles is smaller in tropical and subtropical regions, presenting a more intensive distribution pattern in vascular bundles. The trait variation between subpopulations for the remaining 24 indicators is shown in Figure S5.

**MaizeSPD—phenotype database of stem vascular bundles for NP Population**

Based on the micro-phenotypic data of stem from NP population, the first database for stem micro-phenotypes, MaizeSPD, was established. All data from MaizeSPD were stored and
managed in a MySQL relational database, and a user-friendly web interface was developed to help users search and use the data. MaizeSPD consists of seven data tables: namely information table of maize inbred lines, experimental information table, index table of experimental results, link table of stem CT scanning images, link table of stem images processed by VesselParser, micro-phenotypic indicators list of stem vascular bundles and micro-phenotypic data table of NP population inbred lines. Currently, MaizeSPD has stored 554 pieces of basic information for maize inbred lines, 523 pieces of experimental information, 1008 pieces of CT scanning images and processed images, and 24,192 pieces of stem micro-phenotypic data classified as 48 categories. MaizeSPD is the first database for stem micro-phenotypic information, which lays a foundation for the storage, sharing and improvement of microscopic phenotype data (Video S1).

Significant SNPs obtained by genome-wide association study

In this study, multi-locus random-SNP-effect mixed linear models in R package ‘mrMLM’ (version 4.0) was used to carry out genome-wide association analysis on 30 stem and vascular bundle phenotypic traits. Finally, a total of 1562 significant associated SNPs ($P$-value < 6.4e-7) were identified for target traits. Because these results were a collection of six GWAS methods, the top one most significant SNPs obtained by each method and the SNPs validated by two or more methods were considered as highly significant results. And the reliability of these results could be higher than that of the others. Consequently, 292 highly significant associated SNPs were filtered for all 30 key traits. The detailed statistical results of highly significant SNPs for each trait are shown in Table 1.
Figure 3 The broad-sense heritability ($H^2$) of the investigated 48 phenotypic traits.
Candidate genes co-localized with associated SNPs

All candidate genes were annotated according to the latest maize B73 reference genome (B73 RefGen_v4) available in Ensembl-Plants and NCBI Gene database. In total, 2348 unique candidate genes were annotated by 1562 significant associated SNPs. The number of single-trait-related candidate genes annotated by SNPs listed in multiple methods validated results was 416. For the 416 single-trait-related candidate genes, the NCBI Gene database was used for further annotation, and 294 genes with more detailed functional annotation were obtained (Table S3). Additionally, 84 genes listed both in top one of each method’s results and multiple methods validated results were identified (marked as ‘Y’ in Table 2). Since these genes were not only annotated by the top one SNPs of each method, but also validated by multiple methods, the reliability of these genes was considered to be higher than that of other genes. Among them, the numbers of unique loci associated with each trait were 5 (VBNum), 2 (CSArea), 7 (EZArea), 2 (IZArea), 4 (IZCirR), 2 (IZInsCirR), 3 (IZLALen), 5 (IZSALen), 3 (IZVBAvArea), 4 (IZVBDensity), 7 (IZVBNum), 4 (IZVBVoAvArea), 2 (PZCirR), 1 (PZInsCirR), 3 (PZLALen), 4 (PZVBAvArea), 8 (PZVBDensity), 9 (PZVBNum), 5 (VBAvArea) and 4 (VBDensity). Remarkably, we found a set of genes distributing on 2, 3, 4, 9, 10 chromosomes associated with vascular bundle numbers traits, which involved in the plant signal transduction and stress response; candidate gene distributing on 4 chromosomes associated with vascular bundle area trait, which involved in the gibberellin biosynthesis; several candidate genes distributing on 4, 8 chromosomes associated with vascular bundle distribution density traits, which involved reproductive processes and embryogenesis; and a set of candidate genes distributing on 3, 8 chromosomes associated with epidermis area traits, which encoded enzymes involved in cell wall metabolism.

Pathways enriched by functional enrichment analysis

Functional enrichment analysis was completed to further explore the function of the genes associated with seven dry matter traits. After uploading the candidate gene IDs of all 30 phenotypic traits to PlantRegMap and KOBAS 3.0, a total of 172 GO terms and 8 KEGG pathways (P-value < 0.05) were enriched by these two methods (Table S4 and Figure S6). As mentioned above, 30 traits were clustered into three groups by hierarchical clustering using the Pearson correlation as a distance metric, which as groups I, II and III. And the functional enrichment analysis of candidate genes as each group was conducted, 69 GO BP terms and 4 KEGG pathways (P-value < 0.05) were enriched for 21 phenotypic traits from group I, 27 GO BP terms and 4 KEGG pathways (P-value < 0.05) were enriched for five phenotypic traits from group II, and 87 GO BP terms were enriched for three phenotypic traits from group III. For the group I, geometric and morphological characters of stem cross section and functional zones, and quantitative character of vascular bundles, two GO BP terms ‘negative regulation of reproductive process’ (GO:2000242, P-value = 0.00015), ‘negative regulation of post-embryonic development’ (GO:0048581, P-value = 0.00062) that both containing genes GRMZM2G161913 and GRMZM2G348238 were obtained with the highest significance. In addition, the four significant KEGG pathways ‘Metabolic pathways’ (zma01100, P-value = 0.0170), ‘RNA transport’ (zma03013, P-value = 0.0442), ‘Fatty acid degradation’ (zma00071, P-value = 0.0172) and ‘Tyrosine metabolism’ (zma00350, P-value = 0.0096) were enriched by KOBAS 3.0 (Figure S6A,D). For the group II, related to vascular

Figure 4 The trait variation between subpopulations (TST, NSS, SS and mixed) for the VBNum, VBAvArea, IZVBNum, IZVBVoAvArea, PZVBAvArea and PZVBDensity. The 1–3 figures in the first row represent VBNum, IZVBNum and PZVBDensity, respectively. The 1–3 figures in the second row represent VBAvArea, PZVBAvArea and IZVBVoAvArea, respectively. ** denotes significant differences between subpopulations at P ≤ 0.01 probability level, and * denotes significant differences between subpopulations at P ≤ 0.05.

© 2020 The Authors. Plant Biotechnology Journal published by Society for Experimental Biology and The Association of Applied Biologists and John Wiley & Sons Ltd, 19, 35–50
bundle area characteristics, two GO BP terms ‘nucleobase-containing compound biosynthetic process’ (GO:0034654, P-value = 0.0159) and ‘aromatic compound biosynthetic process’ (GO:0019438, P-value = 0.0186) were obtained with the highest significance. In addition, the four significant KEGG pathways ‘Fructose and mannose metabolism’ (zma00051, P-value = 0.0069), ‘N-Glycan biosynthesis’ (zma00510, P-value = 0.0237), ‘Metabolic pathways’ (zma01100, P-value = 0.0254), ‘Flavone and flavonol biosynthesis’ (zma00944, P-value = 0.0353) and ‘Protein processing in endoplasmic reticulum’ (zma04141, P-value = 0.0443) were enriched by KOBAS 3.0 (Figure S6B,E). For the group III, distribution properties of vascular bundle, two GO BP terms ‘regulation of biological process’ (GO:0050789, P-value = 0.0005) and ‘biological regulation’ (GO:0065007, P-value = 0.00078) that both containing genes GRMZM2G066041, GRMZM2G107101 and GRMZM2G307823 were obtained with a high significance (Figure S6C).

### Discussion

The scope of plant phenotyping has expanded from plant population, single plant, to tissue and cell scales. Apart from the visual traits, internal anatomical traits are equally important; however, few advances have been made in high-performance micro-phenotyping. X-ray micro-CT can perform non-destructive, non-invasive, and three-dimensional visualization and quantification of the internal structure of biological material.
| Traits | Gene   | Description                                      | Chromosome | Position | Alleles   | SNP   | MAF       | Multimethods | Genes listed |
|--------|--------|--------------------------------------------------|------------|----------|-----------|-------|-----------|--------------|--------------|
| CSArea | GRMZM2G002002 | Nucleotidyltransferase                           | 1          | 20699924 | G         | chr1.S_20699924 | 0.0787 | 1         | Y            |              |
| CSArea | GRMZM2G134367 | Nodulin-related protein 1                       | 1          | 20699924 | G         | chr1.S_20699924 | 0.0787 | 1         | Y            |              |
| EZArea | ZEAMMMB73_Zm00001d008913 | Uncharacterized LOC100382589             | 8          | 25128942 | G         | chr8.S_25128942 | 0.451  | 1         | Y            |              |
| EZArea | GRMZM5G12425 | Phospholipase A1 EG1, chloroplast/cytosolic    | 8          | 168955849| G         | chr8.S_168955849| 0.4088 | 2         | Y            |              |
| EZArea | GRMZM2G381473 | Uncharacterized LOC10029339                    | 3          | 57096327 | C         | chr3.S_57096327 | 0.4396 | 2         | Y            |              |
| EZArea | GRMZM2G033644 | Dihydrolipoyllysine-residue acetyltransferase component of pyruvate dehydrogenase complex | 3          | 57096327 | C         | chr3.S_57096327 | 0.4396 | 2         | Y            |              |
| EZArea | GRMZM2G142334 | Putative thioredoxin superfamily protein       | 4          | 150570610| T         | chr4.S_150570610| 0.168  | 1         | Y            |              |
| IZArea | GRMZM2G133475 | Uncharacterized LOC541674                      | 2          | 22358918 | C         | chr2.S_22358918 | 0.0805 | 1         | Y            |              |
| IZArea | GRMZM2G450717 | Peroxidase 52                                   | 2          | 5425011  | A         | chr2.S_5425011  | 0.2651 | 1         | Y            |              |
| IZCirR | GRMZM5G882758 | Pentatricopeptide repeat-containing protein At5g08510 | 9          | 25254192 | C         | chr9.S_25254192| 0.0646 | 1         | Y            |              |
| IZCirR | GRMZM2G021973 | LOC103626187-like pseudogene                   | 5          | 31333610 | T         | chr5.S_31333610| 0.0422 | 1         | Y            |              |
| IZCirR | GRMZM2G390804 | Uncharacterized LOC103626185                  | 5          | 31333610 | T         | chr5.S_31333610| 0.0422 | 1         | Y            |              |
| IZLALen | GRMZM2G145935 | FIP1(V)-like protein                           | 5          | 10079109 | T         | chr5.S_10079109| 0.0976 | 1         | Y            |              |
| IZLALen | GRMZM2G084891, GRMZM5G85064 | DNA-directed RNA polymerase II subunit RP232 | 5          | 17275214 | G         | chr5.S_17275214| 0.2889 | 1         | Y            |              |
| IZLALen | GRMZM2G126447 | Ubiquitin carboxy-terminal hydrolase 15       | 5          | 14798323 | T         | chr5.S_14798323| 0.1562 | 1         | Y            |              |
| IZSALen | GRMZM2G098577 | Uncharacterized LOC100382039                   | 10         | 12970447 | T         | chr10.S_12970447| 0.1785 | 1         | Y            |              |
| IZSALen | ZEAMMMB73_Zm00001d041038 | Uncharacterized LOC103650398                  | 3          | 92384170 | C         | chr9.S_92384170| 0.0739 | 1         | Y            |              |
| IZSALen | ZEAMMMB73_Zm00001d046782 | DNA polymerase eta                             | 9          | 105824379| G         | chr9.S_105824379| 0.4318 | 1         | Y            |              |
| IZSALen | GRMZM2G114680 | Protein MONOCULM 1                             | 9          | 105824379| G         | chr9.S_105824379| 0.4318 | 1         | Y            |              |
| IZSALen | GRMZM5G865319 | Protein Z                                       | 4          | 17762350 | G         | chr5.S_17762350| 0.1024 | 1         | Y            |              |
| IZVBAvea | GRMZM2G167283 | Uncharacterized LOC100273124                   | 3          | 20123209 | G         | chr5.S_20123209| 0.0208 | 2         | Y            |              |
| IZVBAvea | GRMZM5G854666 | Uncharacterized LOC103652335                   | 9          | 76536683 | A         | chr5.S_76536683| 0.2297 | 1         | Y            |              |
| IZVBAvea | GRMZM5G065694 | EREBP-4-like protein                            | 9          | 76536683 | A         | chr5.S_76536683| 0.2297 | 1         | Y            |              |
| IZVBDensity | GRMZM2G107101 | Uncharacterized LOC100147733                  | 8          | 21822571 | A         | chr5.S_21822571| 0.1732 | 1         | Y            |              |
| IZVBDensity | GRMZM2G087416 | Uncharacterized LOC100192922                   | 8          | 21822571 | A         | chr5.S_21822571| 0.1732 | 1         | Y            |              |
| IZVBDensity | GRMZM2G350793 | Probable LRR receptor-like serine/threonine-protein kinase A3g47570 | 1          | 171535280| G         | chr5.S_171535280| 0.0984 | 3         | Y            |              |
| IZVBDensity | GRMZM2G307823, GRMZM2G409627 | NF-X1-type zinc finger protein NF671 | 9          | 479733  | C         | chr5.S_479733 | 0.2178 | 1         | Y            |              |
| IZVBNum | GRMZM2G339562 | Uncharacterized LOC100274353                   | 5          | 23764039 | C         | chr5.S_23764039| 0.0643 | 2         | Y            |              |
| IZVBNum | ZEAMMMB73_Zm00001d013884 | Uncharacterized LOC103626113                  | 5          | 23764039 | C         | chr5.S_23764039| 0.0643 | 2         | Y            |              |
| Traits | Gene             | Description                                      | Chromosome | Position | Alleles | SNP    | MAF      | Multimethods | Genes listed |
|--------|------------------|--------------------------------------------------|------------|----------|---------|--------|---------|--------------|--------------|
|        | GRMZM2G048962    | Anther-specific proline-rich protein APG        | 6          | 122034792| C       | ch6.S,122034792| 0.1076  | 1         | Y            |              |
|        | GRMZM2G155546    | Ribonucleoside-diphosphate reductase small chain | 6          | 122034792| C       | ch6.S,122034792| 0.1076  | 1         | Y            |              |
|        | GRMZM2G168214    | Uncharacterized LOC1003821169                    | 9          | 27177987 | T       | ch9.S,27177987 | 0.0726  | 1         | Y            |              |
|        | GRMZM2G084739    | 60S ribosomal protein L9                         | 9          | 27177987 | T       | ch9.S,27177987 | 0.0726  | 1         | Y            |              |
|        | GRMZM2G322593    | RHOMBOID-like protein 9 chloroplastic            | 4          | 230734500| G       | ch4.S,230734500| 0.2558  | 2         | Y            |              |
|        | GRMZM2G086766    | Tetraspanin-6                                    | 8          | 135351047| A       | ch8.S,135351047| 0.1299  | 1         | Y            |              |
|        | GRMZM2G155546    | Ribonucleoside-diphosphate reductase small chain | 6          | 122034792| C       | ch6.S,122034792| 0.1076  | 1         | Y            |              |
|        | GRMZM2G168214    | Uncharacterized LOC1003821169                    | 9          | 27177987 | T       | ch9.S,27177987 | 0.0726  | 1         | Y            |              |
|        | GRMZM2G084739    | 60S ribosomal protein L9                         | 9          | 27177987 | T       | ch9.S,27177987 | 0.0726  | 1         | Y            |              |
|        | GRMZM2G322593    | RHOMBOID-like protein 9 chloroplastic            | 4          | 230734500| G       | ch4.S,230734500| 0.2558  | 2         | Y            |              |

© 2020 The Authors. Plant Biotechnology Journal published by Society for Experimental Biology and The Association of Applied Biologists and John Wiley & Sons Ltd, 19, 35–50
Table 2

| Chromosome | Description          | Alleles                                                                 |
|------------|----------------------|----------------------------------------------------------------------|
| 1          | VBAvArea             | GRMZM2G057416 Uncharacterized LOC100216812                           |
|            |                      | chr5:1,15897480 0.1082                                              |
| 4          | VBAvArea             | GRMZM2G122503 Benzyl alcohol O-benzoyltransferase                     |
|            |                      | chr4:6743055 T 0.0906                                               |
| 2          | VBAvArea             | GRMZM2G445854 Ent-copalyl diphosphate synthase                        |
|            |                      | chr4:6743055 T 0.0906                                               |
| 1          | VBDensity            | ZEAMMB73_Zm00001d034336 LOC100274439-like pseudogene                 |
|            |                      | chr1:290113928 C 0.1877                                             |
| 3          | VBDensity            | GRMZM2G061695 Gibberellin responsive 1                               |
|            |                      | chr3:223096267 G 0.3556                                             |
| 2          | VBDensity            | GRMZM2G126603 LIN1 protein                                           |
|            |                      | chr3:223096267 G 0.3556                                             |
| 4          | VBDensity            | ZEAMMB73_Zm00001d052115 Probable polyribonucleotide nucleotidyltransferase 1, chloroplastic |
|            |                      | chr4:180199431 C 0.1029                                             |
| 4          | VBNum                | GRMZM2G038108 Activator of 90 kDa heat shock protein ATPase           |
|            |                      | chr4:67144156 T 0.1483                                             |
| 3          | VBNum                | ZEAMMB73_Zm00001d051364 Peptidyl-prolyl cis-trans isomerase CYP20-3, chloroplastic |
|            |                      | chr3:156023882 G 0.2558                                             |
| 4          | VBNum                | GRMZM2G047128 Uncharacterized LOC100277438                          |
|            |                      | chr4:156023882 G 0.2558                                             |

MAF, minor allele frequency. *SNP indicates the significant loci associated with each trait.

*The allele represents the favourable allele.* Leading SNP of each significant locus associated with each trait. Validated by how many methods.

Table 2 Continued

| Traits        | Multimethods | Genes listed |
|---------------|--------------|--------------|
|              | MAF          | top 1        |
|              | 1            | Y            |
|              | 2            | Y            |
|              | 3            | Y            |

with a minimum resolution of 1 µm (Cnudde and Boone, 2013; Landis and Keane, 2010; Zhao et al., 2019). Based on CT images, the image analysis software VesselParser 1.0 developed by du et al. (2016) realized the high-throughput detection of vascular bundle phenotypic traits of stem for the first time. However, it was difficult to segment and analyse the vascular bundle phenotypic traits of mature stem and basal stem, leaving much room for improvement. By previous research, we developed a standard process for stem micro-CT data acquisition and automatic CT image process pipeline to extract vascular bundle traits, which provided the possibility for the microscopic phenotype analysis of large-scale maize accessions. Through VesselParser 4.0 pipeline, contour representations of the slice, functional zones, layers, and vascular bundles provided uniform analysis to output lots of traits, such as geometry-related, morphology-related and distribution-related traits. Compared to other vascular bundles phenotyping methods, VesselParser 4.0 has the following advantages:

(i) ‘zone’ defining provided a basis for the segmentation strategy of stem vascular bundles. There were significant differences in vascular bundles between periphery and inner region, including vascular bundle area, cavity size, CT value and distribution density. ‘Zone’ provides the most suitable classification criteria for the vascular bundles of stem. In ‘inner zone’, vascular bundles are independent of each other and can be separated by simple threshold segmentation. In the periphery zone, vascular bundles are connected by surrounding parenchyma, so more adaptive segmentation strategies should be adopted. (ii) Zone’ reflected the material contents in the stem, structural characteristics of vascular bundle and distribution changes. Based on ‘zone’ image segmentation, quantitative analysis of phenotypic characteristics of vascular bundles in different regions of stem could be achieved, which were more novel and accurate than by manually measuring. The new phenotypic indexes such as zone area, vascular bundle density, might have a better correlation with crop production and will provide new insights into the genetic architecture of vascular system.

According to the information of population structure in previous study (Yang et al., 2011), phenotypic variation of 30 phenotypic indicators between subgroups was compared and obvious differences were identified (Figures 4 and S5). In this study, 1.03- to 9.74-fold variations of vascular bundle traits were detected in a larger sample size of association mapping panel which consisted of 480 inbred lines across whole world. The phenotypic indicators with the most significant differences (P ≤ 0.001) between subpopulations were VBNum, VBAvArea, IZVBNum, IZVBAvArea, PZVBAvArea and PZVBDensity. In addition to the number of vascular bundles, the characteristics of vascular bundle area and vascular bundle density are important references for distinguishing different genotypes. Using VesselParser 4.0, more comprehensively phenotypic information of vascular bundles was captured, thus further expanding our understanding of variations of vascular bundle traits between subgroups. Due to the traditional manual testing method was time-consuming and laborious, only the research of vascular bundle number in the uppermost internode of maize stem was reported (Huang et al., 2016). With greater understanding of the true multivariate and multiscale nature of genotypes will come increased insight into the mechanistic and developmental underpinnings of vascular bundle form and function.

A huge amount of complex, integration of a wide range of image, spectral, environmental data can be generated through by
the high-throughput phenotyping technologies. Thus, the efficient storage, management and retrieval of phenotypic data are becoming the important issues to be considered (Zhao et al., 2019). In recent years, the database and management system of plant phenomics have been reported. 2011, PHENOPSIS DB for Arabidopsis thaliana phenotypic data were established (Fabre et al., 2011); 2014, ClearedLeaves DB, an on open online database was built to store, manage and access cleared leaf images and phenotypic data (Das et al., 2014); 2016, the Leibniz Institute of Plant Genetics and Crop Plant Research and the German Plant Phenotyping Network jointly launched the PGP repository as infrastructure to publish plant phenotypic and genotypic data comprehensively (Arend et al., 2016). However, because of the requirement for accurate identification of microscopic phenotypes for large amount population and a lack of high-throughput and effective micro-phenotyping detection methods, data sets or data management systems for microscopic phenotype information are rarely retrieved. Here, based on the micro-phenotypic data of stem from NP population, the first database for stem micro-phenotypes, MaizeSPD, was established. Currently, MaizeSPD has stored 554 pieces of basic information for maize inbred lines, 523 pieces of experimental information, 1008 pieces of CT scanning images and processed images, and 24 192 pieces of stem micro-phenotypic data classified as 48 categories. MaizeSPD is a successful example of crop microscopic phenotype data storage, management, retrieve and sharing, which lays a foundation in accumulating and improving microscopic phenotype data. In the future, we will continue to expand the information in the database, from stem microscopic data to kernel, leaf and root micro-phenotypic data.

Genomic data play a major role in crop genetic improvements and breeding programmes. However, considerable gains can only be achieved by tightly coupling genomic discoveries to plant phenomics (Cobb et al., 2013). High-throughput and automatic phenotyping facilities in indoor and field environment have developed rapidly over the last 5 years, significantly improving the efficiency and accuracy of crop phenotyping. Large-scale phenotyping has become an important compliment to genome sequencing and identifying genetic regulatory mechanisms (Harfouche et al., 2019; Zhao et al., 2019). In recent years, multomics techniques that combine genomic data with phenotypic data have been applied to crop plants, rapidly decoding the function of a large number of unknown genes and identifying the molecular basis of many agronomic traits (Busemeyer et al., 2013; Muraya et al., 2017; Salvi and Tuberosa, 2015; Wu et al., 2019; Yang et al., 2014a,b, 2015; Zhang et al., 2017). In this study, vascular bundle traits of maize natural population panel containing 480 inbred lines were analysed. Combined with genome-wide association studies (GWASS), a total of 1562 significant single nucleotide polymorphisms (SNPs) were identified for 30 stem micro-phenotypic traits. The number of single-trait-related candidate genes annotated by SNPs listed in multiple methods validated results was 416. For the 416 single-trait-related candidate genes, the NCBI Gene database was used for further annotation, and 294 genes with more detailed functional annotation were obtained. Additionally, 84 genes listed both in top one of each method’s results and multiple methods validated results were identified. Candidate genes identified by GWAS mainly encode enzymes involved in cell wall metabolism, transcription factors, protein kinase and protein related to plant signal transduction and stress response.

Remarkably, we found a set of genes involved in the plant signal transduction and stress response, which associated with vascular bundle numbers traits. The SNP on chromosome 9 significantly associated with MYB transcription factor family genes was found located within the gene model GRMZM2G022686, which encodes a MYB-related protein Myb4. This protein plays important roles in plant dwarf phenotype and increased tolerance to cold and freezing in Arabidopsis and barley (Soltész et al., 2012). And SNP on chromosome 3 significantly associated with MYB transcription factor family genes was also found located within the gene model GRMZM2G348238, which encodes a MYB family transcription factor EFM. This transcription factor acts as a flowering repressor, directly repressing FT expression in a dosage-dependent manner in the leaf vasculature (Yang et al., 2014). Another significant SNP located on chromosome 8 at position 16948568 is contained in the gene region of GRMZM2G042756 that encodes dehydration-responsive element-binding protein 1D, which is a transcription factor played significant roles in responses to biotic and abiotic stresses (Zhang et al., 2012). And the significant SNP located on chromosome 4 is contained in the gene region of Zm00001d051364, which encodes peptidyl-prolyl cis-trans isomerase CYP20-3. Peptidyl-prolyl cis-trans isomerase CYP20-3 is a jasmonate family binding protein, and the jasmonate family of phytohormones plays central roles in plant development and stress acclimation (Dominguez-Solis et al., 2008). For the vascular bundle area trait, we found genes involved in the gibberellin biosynthesis. The significant SNP located on chromosome 4 at position 6743055 is contained in the gene region of GRMZM2G445854 that encodes ent-copaly diphasphate synthase 2, catalysing the conversion of geranyl-geranyl diphsphate to the gibberellin precursor ent-copaly diphasphate (ent-CPP) in responses to biotic and abiotic stresses (Harris et al., 2005; Mafu et al., 2018). A set of candidate genes encoded enzymes involved in cell wall metabolism were identified, which associated with epidermis area traits. The significant SNP located on chromosome 3 at position 57096327 is contained in the gene region of GRMZM2G381473 that encodes UDP-glucuronic acid decarboxylase 4. UDP-glucurionate decarboxylase is the key enzyme involved in the biosynthesis of UDP-α-D-xylene, which is a nucleotide sugar involved in the synthesis of diverse plant cell wall hemicelluloses (xyloglucan, xylan) and minor plant metabolites. What is also interesting is that we found genes involved reproductive processes and embryogenesis, which associated with PZVBDensity traits. The significant SNP located on chromosome 4 at position 6406610 is contained in the gene region of GRMZM2G353267 that encodes 19 kDa zein A30, which is specifically expressed during seed development (Song et al., 2011). Another significant SNP located on chromosome 8 at position 174156791 is contained in the gene region of GRMZM2G339736 that encodes aspartic proteinase PCS1. In Arabidopsis, PCS1, which encodes an aspartic protease, has an important role in determining the fate of cells in embryonic development and in reproduction processes (Ge et al., 2005). Because the anatomical phenotypes of stem vascular bundles are comparatively difficult to obtain and there have been few genetic studies on microscopic traits, the phenotype-associated genes obtained in our study can provide a reference for related research and provide new ideas for exploring the genetic mechanisms of vascular bundle agronomic traits in future studies.
Experimental procedures

Materials, growth conditions and sample collection

Four hundred and eighty maize inbred lines used in this study belonged to the maize natural population described by Yang et al. (2011), which were classified into four subgroups based on population structure Q matrix: Stiff stalk (SS) with 30 lines, non-stiff stalk (NSS) with 133 lines, tropical-subtropical (TST) with 212 lines and an admixed group with 105 lines. The population had a high-density genotype of 1.25 million single nucleotide polymorphism (SNPs) with minor allele frequency (MAF) of > 0.0534 (Liu et al., 2017). The plants were grown in Tongzhou Experimental Station of Beijing Academy of Agriculture and Forestry Sciences in Beijing, China (116.68°E, 39.69°N). Sowing took place on 28 April 2018. Each inbred line was planted in four-row plot, with eight plants each row. Each row was 2.1 m long and 60 cm between rows. The third internodes of three plants for each inbred line were collected at the silking period (73 days after sowing) for later research.

Standard scanning protocol and CT image reconstruction

The standardized procedure of maize microdata acquisition was constructed to ensure the reliability and consistency of image acquisition, and to provide a solution for large-population and high-precision scanning imaging of crop. First, we developed a sample preparation protocol for maize stem. A motor electric cutting machine (Bosch stone cutting machine, GDM13-34) was firstly used to cut the third internode of maize stem into a series of 0.5–1.0 cm segments, since it was strenuous and fail-prone by manual cutting. The sample segments were soak in FAA solution (90:5:5 v/v/v, 70% ethanol:100% formaldehyde:100% acetic acid) immediately. After the FAA fixation, samples were performed the sequential ethanol gradient dehydration in bath (i.e. 70%, 95% and 100%) and set the processing time of each ethanol gradient as 30 min. Next, samples were transferred to tertiary butyl alcohol and soaked for 24 h, and then froze samples at −80 °C for 24 h. Finally, frozen samples were placed in the freeze-dryer (LGJ-10E, China) and freeze-dried for 3 h in batch. According to the micro-CT scanning protocol introduced by Zhang (Zhang et al., 2018), dried stem samples were scanned by Skyscan 1172, and the unified scanning parameter was set as: 40 kV/250 μA, the imaging pixel sizes as 13.55 μm, 2K scanning mode (2000 × 2000 pixels). Finally, we defined much stricter reconstruction parameters to guarantee the consistence and quantification of imaging quality. According to the linear absorbance coefficient for X-ray of various materials, the Hounsfield (HU) values of air and water are, respectively, −1000 and 0. For different maize variety and growth stages, we found out that HU values of maize stem distributed in a wide range from 0 to 7000. To provide a standard for quantification and evaluation, we defined a wider value ranges covered whole HU ranges of plant materials, that is [−1000, 9240], to transform raw data (16 grey level) into an 8-bit grey-level image with a value range [0, 255] (Figure S7).

Image segmentation and analysis strategies

Here, we designed a fully automated image analysis pipeline based C++ and OpenCV, named as VesselParser 4.0. This image analysis pipeline was summarized as a flow chart in figure 1. As an important innovation, the transverse structures of maize stem were divided into three zones with physiological significance based on the material distribution and relationship, that is epidermis zone, periphery zone and inner zone. From imaging viewpoint, these zones were demonstrated different pixel intensity (matching to CT Hounsfield value) and pixel connectivity.

In our knowledge, the physiological zones of maize stem are difficult to be accurately measured by manual work and visual investigation owing to the boundary ambiguity. Here, we first defined and detected these zones in the pixel level of image. Moreover, these zones could be described using simpler pattern description, such as boundary contours. Based on an object-oriented strategy, we built a three-layer structure (i.e. maize stem, zones and vascular bundles, respectively) to represent the maize stem slice for all maize cultivars. For each layer, the image analysis scheme of vascular bundles was more specific and robust according to the zone properties. As a result, more valuable traits related with three-layer structures could be, respectively, detected. For example, epidermis region can be almost regarded as the epidermis zone; thus once the epidermis zone can be precisely detected based CT image, the epidermis thickness is easy to be measured, and as everyone knows, it is difficult to be done by manual work.

Data detection and selection

The pre-processing of phenotypic data involved outlier detection and trait reproducibility assessments. We first adopted Grubbs’ test (Grubbs, 1950) to evaluate the repeatability of phenotypic traits based on the assumption of normally distributed phenotypic data points for repeated measurements on replicated plants with a single genotype for each trait. Grubbs test results with a P-value < 0.01 were considered to be outliers. The frequency of outliers in the reproducibility trait should be less than the number of random occurrences. Then, the Pearson correlation coefficient was used to detect sample outliers. A sample with r < 0.8 among all the other samples was identified as an outlier. Next, multiple linear regression (MLR) was implemented to reduce the correlation among exploratory variables. As a general rule, we considered VIF > 5 as a cut-off value for the high multi-collinearity problem. We used the VIF function in the R software package fmsb to calculate VIF. The ‘lm’ function of the R package lmer was used to construct the MLR model. Finally, the variables that remained after the stepwise regression were used for following analysis.

Clustering analysis, ANOVA and heritability analysis

We conducted unsupervised hierarchical clustering analysis using Pearson’s correlation as a distance metric to organize 48 trait data into meaningful structures. In clustering, it considers each trait as a random variable with n observations and measures the similarity between the two traits by calculating the linear relationship between the distributions of the two corresponding random variables (Jiang and Lu, 2007). And the analysis of variance (ANOVA) was carried out to find out whether differences exist between different subpopulations means. These statistical analyses were conducted using SPSS Statistics 22 (2016) software (IBM, Armonk, NY, USA).

Heritability refers to the percentage of genetic variation (Vg) that accounts for the total variation of the phenotype, generally denoted by $H^2$. It can be used to compare the relationship between genetic ($\sigma^2_e$) and environmental ($\sigma^2_e$) factors for a specific phenotypic variation (Vp). Heritability ($H^2$) was calculated for each trait as follows:
The above analysis was performed in ASReml-R v.3.0 by using the ‘asreml’ function of R package asreml (David, 2009).
Genotype data were obtained from Professor Yan Jianbing’s laboratory of Huazhong Agricultural University (download URL: www.maizego.org/Resources.html). After quality control, 779 855 SNPs with minimum allele frequency (MAF) > 0.05 and call rate > 0.01 were used in our study. Population structure was estimated using the STRUCTURE program version 2.3.4 (Hubisz et al., 2009) based on 480 maize inbred lines, and 779 855 SNPs were used to estimate the relative kinship by TASSEL 5 (Bradbury et al., 2007). A multi-locus random-SNP-effect mixed linear model tool for GWAS (R package ‘mrMLM’ version 4.0) (Zhang et al., 2019) was used on the 30 phenotypic traits separately to test the statistical association between phenotypes and genotypes. Population structure and relative kinship were taken into account in the model, and six methods (mrMLM, FASTmrMLM, FASTmrEMMA, ISIS EM-BLASSO, pLARmEB and pKWmEB) included in the function ‘mrMLM’ were used in our study. For the first step, the criterion of P-value was set as 6.4e-7 (P ≤ 0.5/N, where N is the total number of genome-wide SNPs). And then the default P-value 0.0002 was used as the filter threshold for the second step to declare significance of SNPs associated with a given trait. The common results obtained by all methods were regarded as significant SNPs associated with phenotypic traits, and the overlapped loci of multiple methods were considered to be more reliable results. All candidate genes were annotated according to the latest maize B73 reference genome (B73 RefGen_v4) available in EnsemblPlants (http://plants.ensembl.org/Zea_mays/Info/Index) and NCBI Gene database (https://www.ncbi.nlm.nih.gov/gene).

Functional and network analysis
Pathway enrichment analysis was performed by PlantRegMap (Jin et al., 2015, 2017) and KOBAS 3.0 (Wu et al., 2006; Xie et al., 2011). The input data were consisted of all significant genes annotated by SNPs in coding regions for each phenotypic trait. Gene Ontology (GO) (Ashburner et al., 2000) terms with P-value < 0.05 were identified as significant results. Besides, the significant interactions between the genes and their related phenotypic traits were visualized using Cytoscape v.3.7.2 (National Institute of General Medical Sciences, Bethesda, MD, USA).

Acknowledgements
This work was supported by the Construction of Collaborative Innovation Centre of Beijing Academy and Forestry Science (Collaborative Innovation Centre of Crop Phenomics, KCX201917), the National Nature Science Foundation of China (No. 31801254 and 31671577), Beijing Academy of Agricultural and Forestry Sciences Grant (No. PT2020-24; No. KCX20180423) and China Agriculture Research System (CARS-02). We thank the Maize Research Center department of the Beijing Academy of Agriculture and Forestry Sciences for preparing the seed and planting for the trial. We thank Pro. Jianbing Yan, from Huazhong Agricultural University, Pro. Xiaohong Yang, from the National Maize Improvement Center of China, China Agricultural University, for providing seeds of the maize inbred lines. We also thank Chinese Academy of Agricultural Sciences (CAAS) for providing seeds of the maize inbred lines.

Conflict of interest
The authors declare that they have no conflicts of interest.

Author contributions
Y Zhang drafted and revised the manuscript. X Guo and C Zhao proposed the conceptualization of this study and reviewed the manuscript. J Du and J Wang edited the manuscript. Y Zhang, Y Zhao, X Lu, W Wen, S Gu, J Fan, C Wang, Y Wang, S Wu and S Liao performed field experiments and collected image data. J Du developed the image processing pipeline. Y Zhang and J Wang implemented the statistical analysis and GWAS work. J Wang implements MaizeSPD construction. Y Zhang, J Du and J Wang analysed and interpreted the results. All authors read and approved the final manuscript.

References
Arend, D., Junker, A., Scholz, U., Schüler, D., Wyle, J. and Lange, M. (2016) PGP repository: a plant phenomics and genomics data publication infrastructure. Database, 2016, baw033.
Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P. et al. (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat. Genet. 25, 25–29.
Bai, X., Wu, B. and Xing, Y. (2012) Yield-related QTLs and their applications in rice genetic improvement. J. Integr. Plant Biol. 54, 300–311.
Bradbury, P.J., Zhang, Z., Kroon, D.E., Casstevens, T.M., Ramdoss, Y. and Buckler, E.S. (2007) TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics, 23, 2633–2635.
Buckler, E.S., Holland, J.B., Bradbury, P.J., Acharya, C.B., Brown, P.J., Browne, C. et al. (2009) The genetic architecture of maize flowering time. Science, 325(5941), 714–718.
Busemeyer, L., Ruckelshausen, A., Möller, K., Melchinger, A.E., Alheit, K.V., Maurer, H.P., Hahn, V. et al. (2013) Precision phenotyping of biomass accumulation in triticale reveals temporal genetic patterns of regulation. Sci. Rep. 3, 2442.
Chen, Y. and Liübberstedt, T. (2010) Molecular basis of trait correlations. Trends Plant Sci. 15, 454–464.
Chen, S.-Q., Wang, Y., Zhang, Z.-H., Wang, L.-X., Hu, Z.-L., Li, P., Zhu, L.-H. et al. (2004) Genetic dissection of vascular bundle systems in peduncle and panicle characters in rice (Oryza sativa L.) by means of RFLP markers. J. Plant Biol. Res. 22(1), 15–21.
Chen, J., Zhang, F., Zhao, C., Lv, G., Sun, C., Pan, Y., Guo, X. et al. (2019) Genome-wide association study of six quality traits reveals the association of the TaRPP13L1 gene with flour colour in Chinese bread wheat. Plant Biotechnol. J. 17, 2106–2122.
Crudelle, V. and Boone, M. (2013) High-resolution X-ray computed tomography in geosciences: a review of the current technology and applications. Earth Sci. Rev. 123, 1–17.
Cobb, J.N., Declerck, G., Greenberg, A., Clark, R. and McCouch, S. (2013) Next-generation phenotyping: requirements and strategies for enhancing our understanding of genotype-phenotype relationships and its relevance to crop improvement. Theor. Appl. Genet. 126, 867–887.
Cui, K.H., Peng, S.B., Xing, Y.Z., Yu, S.B., Xu, C.G. and Zhang, Q. (2003) Molecular dissection of the genetic relationships of source, sink and transport tissue with yield traits in rice. Theor. Appl. Genet. 106(4), 649–658.
Das, A., Bucksch, A., Price, C.A. and Wertz, J.S. (2014) ClearedLeavesDB: an online database of cleared plant leaf images. Plant Methods, 10, 8.
David, B. (2009) asreml: asreml fits the linear mixed model. R package version 3.0. www.vsni.co.uk
Del’Acqua, M., Gatti, D.M., Pea, G., Cattonaro, F., Coppens, F., Magris, G., Hiaing, A.L. et al. (2015) Genetic properties of the MAGIC maize population: a new platform for high definition QTL mapping in Zea mays. Genome Biol. 16, 167.
Huang, C., Chen, Q., Xu, G., Xu, D., Tian, J. and Tian, F. (2016) Identification et al.
Jin, J.P., He, K., Tang, X., Li, Z., Lv, L., Zhao, Y., Luo, J.C. et al. (2016) An effective hierarchical clustering methods for
Effective hierarchical clustering methods for plant and ear height in maize. Plant J. 86(5), 391–402.
Li, X., Zhou, Z., Ding, J., Wu, Y., Zhou, B., Wang, R., Ma, J. et al. (2016b) Combined linkage and association mapping reveals QTl and candidate genes for plant and ear height in maize. Front. Plant Sci. 7, 833.
Liu, H., Wang, X., Warburton, M.L., Wen, W., Jin, M., Deng, M. and Liu, J. (2015) Genomic, transcriptomic, and phenomic variation reveals the complex adaptation of modern maize breeding. Molecular Plant, 8, 871–884.
Liu, H., Xue, L., Niu, L., Xiao, Y., Chen, L., Liu, J., Wang, X. et al. (2017) Distant eQTLs and non-coding sequences play critical roles in regulating gene expression and quantitative trait variation in maize. Molecular Plant, 6, 414–426.
Mafu, S., Ding, Y., Murphy, K.M., Yaacobbi, O., Addison, J.B., Wang, Q., Shen, Z. et al. (2018) Discovery, biosynthesis and stress-related accumulation of dolabradiene-derived defenses in maize. Plant Physiol. 176(4), 2677–2690.
McConnell, J.R., Emery, J., Eshed, Y., Bao, N., Bowman, J. and Barton, M.K. (2001) Role of PHABULOSA and PHAVOLUTA in determining radial patterning in shoots. Nature, 411, 709–713.
Muraya, M.M., Chu, J., Zhao, Y., Junker, A., Klukas, C., Reif, J.C. and Altman, T. (2017) Genetic variation of growth dynamics in maize (Zea mays L.) revealed through automated non-invasive phenotyping. Plant J. 89, 366–380.
Nátróv, Z. (1985) Wheat stem vascular size in relation to kernel number and weight. Acta Univ. Agri. Brno Facult. Agron. 33(3), 643–646.
Nátróv, Z. (1991) Anatomical characteristics of the uppermost internode of winter wheat genotypes differing in stem length. Biol. Plant. 33, 491–494.
Pfeifer, J.A., Romay, M.C., Gore, M.A., Flint-Garcia, S.A., Zhang, Z., Millard, M.J., Gardner, C.A. et al. (2014) The genetic architecture of maize height. Genetics, 196(4), 1337–1356.
Peterson, D.M., Housley, T.L. and Luk, T.M. (1982) Oat stem vascular size in relation to kernel number and weight. II. Field environment. Crop. Sci. 22, 274–278.
Qi, J., Qian, Q., Bu, Q., Li, S., Chen, Q., Sun, J., Liang, W. et al. (2008) Mutation of the rice Narrow leaf1 gene, which encodes a novel protein, affects vein patterning and polar auxin transport. Plant Physiol. 147, 1947–1959.
Riedelsheimer, C., Lisek, J., Czedik-Eysenberg, A., Sulpice, R., Flis, A., Grieder, C., Altmann, T. et al. (2012) Genome-wide association mapping of leaf metabolic profiles for dissecting complex traits in maize. Proc. Natl Acad. Sci. USA, 109(23), 8872–8877.
Salvi, S. and Tuberosa, R. (2015) The crop QTLome comes of age. Curr. Opin. Biotechnol. 32, 179–185.
Schnable, P.S., Ware, D., Fulton, R.S., Stein, J.C., Wei, F., Pasternak, S. et al. (2009) The B73 maize genome: complexity, diversity, and dynamics. Science, 326(5956), 1122–1125.
Soltész, A., Vágújfalvi, A., Rizza, F., Kerepesi, I., Galiba, G., Cattivelli, L., Coraggio, I. et al. (2012) The rice Osmbyb gene enhances tolerance to frost and improves germination under unfavourable conditions in transgenic barley plants. Plant Genet. 53(2), 133–143.
Song, R., Llaca, V., Linton, E. and Messing, J. (2011) Sequence, regulation, and evolution of the maize 22-kD α zein gene family. Genome Res. 11(11), 1817–1825.
Terao, T., Nagata, K., Morino, K. and Hirose, T. (2010) A gene controlling the formation and hence is responsible to increase the harvest index and grain yield in rice. Theor. Appl. Genet. 120, 875–893.
USDA FAS (2013) USDA FAS. Circular Series: World Cultural Production. http://www.fas.usda.gov/psdonline/circularsproduction/pdf

© 2020 The Authors. Plant Biotechnology Journal published by Society for Experimental Biology and The Association of Applied Biologists and John Wiley & Sons Ltd, 19, 35–50.
Van Inghelbandt, D., Melchinger, A.E., Martinková, J.P. and Stich, B. (2012) Genome-wide association mapping of flowering time and northern corn leaf blight (Setosphaeria turcica) resistance in a vast commercial maize germplasm set. *BMC Plant Biol.* **12**, 56.

Wang, N., Li, F., Wang, Z., Zhou, Y. and Shi, Z. (2011) Progress in study on vascular bundles characters of corn. *J. Maize Sci.* **19**(3), 148–152.

Wang, X., Zhang, R., Song, W., Han, L., Liu, X., Sun, X., Luo, M. et al. (2019) Dynamic plant height QTL revealed in maize through remote sensing phenotyping using a high throughput unmanned aerial vehicle (UAV). *Sci. Rep.* **9**, 3458.

Weng, J., Xie, C., Hao, Z., Wang, J., Liu, C., Li, M., Zhang, D. et al. (2011) Genome-wide association study identifies candidate genes that affect plant height in Chinese elite maize (*Zea mays L.*) inbred lines. *PLoS One* **6**(12), e29229.

Wu, J., Mao, X., Cai, T., Luo, J. and Wei, L. (2006) KOBAS server: a web-based platform for automated annotation and pathway identification. *Nucleic Acids Res.* **34**, W720–W724.

Wu, D., Guo, Z., Ye, J., Feng, H., Liu, J., Chen, G., Zheng, J. et al. (2019) Combining high-throughput micro-CT-RGB phenotyping and genome-wide association study to dissect the genetic architecture of tiller growth in rice. *J. Exp. Bot.* **70**(2), 545–561.

Xiao, Y., Liu, H., Wu, L., Warburton, M. and Yan, J. (2017) Genome-wide association studies in maize: praise and stargaze. *Mol. Plant* **10**(3), 359–374.

Xie, C., Mao, X., Huang, J., Ding, Y., Wu, J., Dong, S., Kong, L. et al. (2011) KOBAS 2.0: a web server for annotation and identification of enriched pathways and diseases. *Nucleic Acids Res.* **39**, W316–W322.

Yan, Y., Shen, L., Chen, Y., Bao, S., Thong, Z. and Yu, H. (2014) A MYB-domain protein EFM mediates flowering responses to environmental cues in *Arabidopsis*. *Dev. Cell* **30**(4), 437–448.

Yang, X., Gao, S., Xu, S., Zhang, Z., Prasanna, B.M., Li, L., Li, J. et al. (2011) Characterization of a global germplasm collection and its potential utility for analysis of complex quantitative traits in maize. *Mol. Breed.* **28**, 511–526.

Yang, Q., Li, Z., Li, W., Ku, L., Wang, C., Ye, J., Li, K. et al. (2013a) CACTA-like transposable element in ZmCCT attenuated photoperiod sensitivity and accelerated the postdomestication spread of maize. *Proc. Natl Acad. Sci. USA* **110**(42), 16969–16974.

Yang, W., Duan, L., Chen, G., Xiong, L. and Liu, Q. (2013b) Plant phenomics and high-throughput phenotyping: accelerating rice functional genomics using multidisciplinary approaches. *Curr. Opin. Plant Biol.* **16**, 180–187.

Yang, N., Lu, Y., Yang, X., Huang, J., Zhou, Y., Ali, F., Wen, W. et al. (2014a) Genome wide association studies using a new nonparametric model reveal the genetic architecture of 17 agronomic traits in an enlarged maize association panel. *PLoS Genet.* **10**(9), e1004573.

Yang, W., Guo, Z., Huang, C., Duan, L., Chen, G., Jiang, N., Fang, W. et al. (2014b) 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.

Zhai, L., Zheng, T., Wang, X., Wang, Y., Chen, K., Wang, S., Wang, Y. et al. (2018) QTL mapping and candidate gene analysis of peduncle vascular bundle related traits in rice by genome-wide association study. *Rice* **11**, 13.

Zhang, J., Ji, H., Zhang, D., Wang, M. and Cao, Y. (2012) Research progress of DREB gene in plant. *J. Shanxi Agric. Sci.* **40**(1), 79–83.

Zhang, Y., Legay, S., Barrière, Y., Mechlin, V. and Legland, D. (2013) Color quantification of stained maize stem section describes lignin spatial distribution within the whole stem. *J. Sci. Food Agric.* **61**, 3186–3192.

Zhao, C., Zhang, Y., Du, J., Guo, X., Wen, W., Gu, S., Wang, J. et al. (2019) Crop phenomics: current status and perspectives. *Front. Plant Sci.* **10**, 714.

Zhang, Y., Ma, L.M., Pan, X.D., Wang, J.L., Guo, X.Y. and Du, J.J. (2018) Micro-scale phenotyping techniques of maize vascular bundles based on X-ray microcomputed tomography. *JoVE*, **14**, e58501.

Zhang, Y.W., Li, P., Ren, W.L., Ni, Y.L. and Zhang, Y.M. (2019) *mrMLM*: Multi-locus random-SNP-effect mixed linear model tools for genome-wide association study. *R* package version 4.0. https://CRAN.R-project.org/package=mrMLM.GUI

Supporting information

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

**Figure S1** The screenshots of VesselParser 4.0 software function modules, including data management module, method parameters module, phenotyping computation module, and statistic analysis module.

**Figure S2** Phenotypic distribution of 48 items phenotypic traits in the third internode of maize stem in NP population.

**Figure S3** The clustering analysis of 48 phenotypic traits conducted with hierarchical clustering using the Pearson correlation as a distance metric.

**Figure S4** Pearson correlation of 48 micro-phenotypic traits of the basal third internode of maize stem in NP population.

**Figure S5** The traits variation between subpopulations (TST, NSS, SS, and Mixed) for the remaining 24 indicators.

**Figure S6** Functional enrichment results of all candidate genes associated with phenotypic traits.

**Figure S7** The flowchart of high-throughput micro-phenotype analysis of maize stem.

**Table S1** The 48 traits of stem description and abbreviation.

**Table S2** The phenotypic variations for 48 traits of vascular bundle in the third internode of maize stem among natural population.

**Table S3** 294 candidate genes with detailed functional annotation by the NCBI Gene database.

**Table S4** GO terms and KEGG pathway enriched by the genes associated with 30 phenotypic traits by PlantRegMap and KOBAS 3.0.

**Video S1** Video introduction of MaizeSPD, the phenotype database of stem vascular bundles for NP population.