

Large-scale metabolite quantitative trait locus analysis provides new insights for high-quality maize improvement

Kun Li^{1,†}, Weiwei Wen^{1,2,3,†}, Saleh Alseekh^{2,4}, Xiaohong Yang⁵, Huan Guo¹, Wenqiang Li¹, Luxi Wang¹, Qingchun Pan¹, Wei Zhan¹, Jie Liu¹, Yanhua Li¹, Xiao Wu¹, Yariv Brotman⁶, Lothar Willmitzer⁴, Jiansheng Li^{5,*}, Alisdair R. Fernie^{2,4,*} and Jianbing Yan^{1,*}

¹National Key Laboratory of Crop Genetic Improvement, Huazhong Agricultural University, Shizishan Lu 1, 430070, Hongshan, Wuhan, China,

²Max Planck Institute of Molecular Plant Physiology, Am Mühlenberg 1, 14476, Potsdam-Golm, Germany,

³Key Laboratory of Horticultural Plant Biology, Ministry of Education, Huazhong Agricultural University, Shizishan Lu 1, 430070, Hongshan, Wuhan, China,

⁴Centre of Plant System Biology and Biotechnology, 4000, Plovdiv, Bulgaria,

⁵Beijing Key Laboratory of Crop Genetic Improvement, National Maize Improvement Center of China, China Agricultural University, West Yuanmingyuan Lu 2, 100193, Haidian, Beijing, China, and

⁶Department of Life Sciences, Ben-Gurion University of the Negev, Beersheba, Israel

Received 22 December 2018; revised 27 February 2019; accepted 11 March 2019.

*For correspondence (e-mails yjianbing@mail.hzau.edu.cn; fernie@mpimp-golm.mpg.de; lijiansheng@cau.edu.cn).

†These authors contributed equally to this paper.

SUMMARY

It is generally recognized that many favorable genes which were lost during domestication, including those related to both nutritional value and stress resistance, remain hidden in wild relatives. To uncover such genes in teosinte, an ancestor of maize, we conducted metabolite profiling in a BC₂F₇ population generated from a cross between the maize wild relative (*Zea mays* ssp. *mexicana*) and maize inbred line Mo17. In total, 65 primary metabolites were quantified in four tissues (seedling-stage leaf, grouting-stage leaf, young kernel and mature kernel) with clear tissue-specific patterns emerging. Three hundred and fifty quantitative trait loci (QTLs) for these metabolites were obtained, which were distributed unevenly across the genome and included two QTL hotspots. Metabolite concentrations frequently increased in the presence of alleles from the teosinte genome while the opposite was observed for grain yield and shape trait QTLs. Combination of the multi-tissue transcriptome and metabolome data provided considerable insight into the metabolic variations between maize and its wild relatives. This study thus identifies favorable genes hidden in the wild relative which should allow us to balance high yield and quality in future modern crop breeding programs.

Keywords: maize, teosinte, primary metabolism, genetic basis, quantitative trait locus.

INTRODUCTION

Maize (*Zea mays* ssp. *mays*) plays a vital role in human nutrition and energy supply. It is also one of the most economically important grain crops for feed, biofuel and industrial raw materials, with an approximately eight-fold increase in yield achieved over the past century (Duvick, 2005). Considerable evidence suggests that maize was domesticated from its progenitor Balsas teosinte (*Zea mays* ssp. *parviglumis*) in mid- to lowland regions of southern Mexico approximately 9000 years ago (Matsuoka *et al.*, 2002; Piperno *et al.*, 2009; Hufford *et al.*, 2012). Maize and teosinte differ remarkably in plant morphology, yet only about 1200 genes were selected during domestication (Wright *et al.*, 2005). Furthermore, the maize transcriptome

has been substantially altered by domestication and improvement processes, with hundreds to thousands of genes having different expression patterns and co-expression profiles between cultivated maize and teosinte (Wright *et al.*, 2005; Hufford *et al.*, 2012; Swanson-Wagner *et al.*, 2012) and among tropical and temperate maize (Liu *et al.*, 2015). The targeted genes were functionally enriched in biotic stress responses, which may reflect both the effects of inbreeding and changes in natural conditions from the place of origin (Swanson-Wagner *et al.*, 2012). Genetic diversity was reduced during domestication, mainly because of directional artificial selection, natural selection and genetic drift. It has been reported that introgression of teosinte into cultivated maize could improve maize

germplasm, resulting in plants that are more resistant to insects, pathogens and abiotic stress (Takahashi *et al.*, 2012).

Considerable progress has been made toward understanding the differences between maize and teosinte, with many genes having been identified as responsible for changes in maize plant architecture and phenology, including *teosinte glume architecture1 (tga1)* (Wang *et al.*, 2005), *teosinte branched1 (tb1)* (Doebley *et al.*, 1997), *barren stalk1 (ba1)* (Gallavotti *et al.*, 2004), *ramosa1 (ra1)* (Vollbrecht *et al.*, 2005), *zfl1* (Bomblies and Doebley, 2006) and *zfl2* (Bomblies and Doebley, 2006). Beyond visible differences, some physiological traits have also been studied, including the levels of phenolics and antioxidant activities (Zavala-López *et al.*, 2018), insect and pathogen resistance (Lange *et al.*, 2014), anti-herbivore defenses (Maag *et al.*, 2015) and resistance to southern leaf blight (Lennon *et al.*, 2017) gray leaf spot (Lennon *et al.*, 2016).

All these differences were determined to be the consequence of either genome diversity and/or transcriptome regulation. Recently, several studies have demonstrated that some specific nutritional and flavor qualities are significantly reduced in modern cultivars compared with their wild progenitors, with the clearest evidence of this being in tomato (Tieman *et al.*, 2017). It was also recently demonstrated that teosinte has a higher protein and lipid content than either the maize landraces or the inbred lines (Flint-Garcia *et al.*, 2009). Furthermore, the difference in maize and teosinte kernel components may be reflected in the metabolite and starch content and eventually influence maize starch characteristics (such as starch viscosity). Thus, teosinte may be useful for promoting maize breeding by reintroducing the favorable exotic alleles lost during the domestication and improvement processes of maize breeding (Zamir, 2001; Takahashi *et al.*, 2012).

Zea mays ssp. *mexicana* (hereafter *mexicana*) is a wild annual grass native to high altitudes in northern and central Mexico and is a close wild relative of cultivated maize. Compared with cultivated maize, *mexicana* has a stronger growth capacity, higher kernel protein content and more dominant resistance to many fungal diseases (Fang *et al.*, 2012). Thus, it is a very important genetic material for maize improvement. Quantitative trait locus (QTL) analysis can help detect genes/loci in wild species that may improve yield- or quality-related traits in elite varieties. Several studies have been reported concerning introgression of favorable traits from wild species into cultigens. Plants overexpressing the *mexicana* allele *ZmmICE1* show more resistance to freezing stress (Lu *et al.*, 2017) and the early flowering allele *ZEA CENTRORADIALIS 8 (ZCN8)*, which originated from *mexicana*, introgressed into maize and contributed to the adaptation of maize to northern high latitudes (Guo *et al.*, 2018). In tomato, the *Tm-2^a* allele

from *Solanum peruvianum* was introgressed into elite cultivars, thus conferring tomato mosaic virus resistance to the cultivars (Verlaan *et al.*, 2013).

Metabolomics has been developed into a powerful tool for evaluating phenotypic variance within broad genetic populations (Luo, 2015; Fernie and Tohge, 2017; Wen *et al.*, 2018). In an attempt to connect the gap between genomes and end-phenotypes, the analysis of differences in metabolite abundance has also proved to be an efficient strategy (Obata *et al.*, 2015). Metabolomics has, furthermore, provided us with considerable insight into both genetic and biochemical regulation of metabolism, with studies including the comparison of cultivated maize and maize landraces (Venkatesh *et al.*, 2016), the study of a maize nested association mapping population (Zhang *et al.*, 2015) and the study of metabolite QTLs in recombinant inbred lines (Wen *et al.*, 2015) or diverse association mapping populations (Wen *et al.*, 2018) all proving highly instructive. We have a relatively clear knowledge and understanding of the difference between maize and teosinte at the levels of genetic diversity and transcriptional regulation (Hufford *et al.*, 2012; Swanson-Wagner *et al.*, 2012); however, we have very limited knowledge concerning the global metabolic differences between the species.

The generation of introgression populations by crossing wild relatives and modern elite cultivars is a good strategy for identifying the hidden genes in wild species and has been widely used in maize and other crops (Schauer *et al.*, 2006; Briggs *et al.*, 2007; Liu *et al.*, 2016; Zhu *et al.*, 2018). Here, a previously developed introgression population created between the elite inbred line Mo17 and teosinte (*mexicana*) (Yang *et al.*, 2017) was used to dissect the hidden genes/loci affecting metabolite abundance. A total of 65 primary metabolites were measured using a GC-MS platform in four different tissues. The QTL mapping analysis of these metabolites provided us an alternative understanding of the genetic regulation of primary metabolism and yield-related traits. It demonstrated that yield was negatively associated with most of the metabolic traits, suggesting that a balance needs to be found between yield and enhanced metabolite levels. These results provide a framework which will allow for the recovery of nutritionally important components via knowledge-driven molecular breeding.

RESULTS

Variation of primary metabolites in multiple tissues of a teosinte–maize BC₂F₇ population

The BC₂F₇ introgression population between Mo17 and teosinte (*mexicana*; hereafter, TM population) contained 191 lines which were analyzed by GC-MS to determine the levels of primary metabolites from four different tissues (i.e. SL, seedling leaf; ML, mature leaf; YK, young kernel;

MK, mature kernel). A total of 65 primary metabolites (62, 61, 56 and 50 in SL, ML, YK and MK tissues, respectively) with known chemical structure were detected. These were classified into five categories: 22 amino acids (20 protein amino acids and 2 non-protein amino acids), 21 organic acids, 12 sugars, 4 amines and 6 miscellaneous metabolites classified hereafter as ‘others’. All 65 metabolites were detected in at least two tissues and 40 were detected in all four tissues (Table S1 in the online Supporting Information). The dramatic difference in morphology between the elite inbred line Mo17 and its collateral ancestor – teosinte – was mirrored by surprising accumulation patterns of primary metabolites. Approximately 60% (134/229) of primary metabolites displayed a more than 10-fold change across the TM population (Figures S1–S4). Larger variation of primary metabolites within the population was observed in the more mature tissues (leaf and kernel) compared with the younger tissues (Figure 1a). The pair-wise correlation

between metabolites from the same tissue is much stronger than that from different tissues (Figure 1b, Table S2), which is in agreement with our previous findings (Wen *et al.*, 2014, 2015). Subsequent cluster analyses were conducted within the data obtained for each independent tissue and in the combined dataset (Figures 1c and S5). Perhaps unsurprisingly, within the same tissue, metabolites of the same compound class are more tightly linked with each other.

Mode of inheritance of genome-wide loci for primary metabolite variance in multiple tissues

Based on the constructed linkage map (Yang *et al.*, 2017) and phenotypic value (primary metabolites, kernel shape and compositions) measured in the population, QTL analysis was performed. In total, 350 QTLs (97 in SL, 68 in ML, 98 in YK and 87 in MK) were detected across the whole genome. The length of the QTL one-LOD support

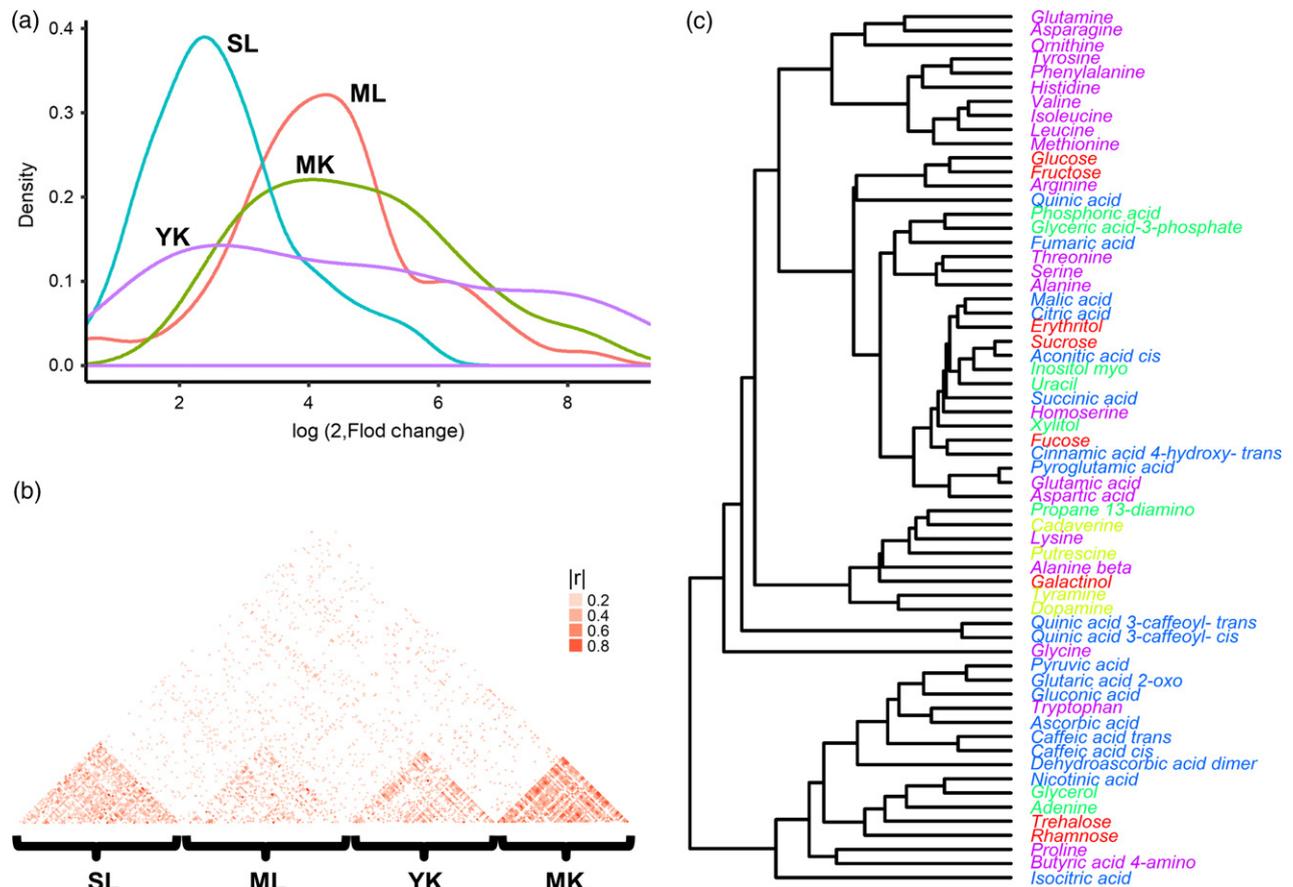


Figure 1. Summary of primary metabolic variation detected in different maize tissues.

(a) Fold changes distribution of each metabolite in different tissues [SL, seedling leaf; ML, mature leaf, i.e. ear leaf at 15 days after pollination (DAP); YK, young kernel, i.e. kernel at 15 DAP; MK, mature kernel].

(b) The Pearson correlation coefficient matrix (under triangle) between pair-wise primary metabolites measured in the same or different tissues. Absolute values of correlations are shaded in red. The strength of the correlation is indicated by red color saturation.

(c) Hierarchical clustering analysis of primary metabolites measured in the seedling leaf stage. Those metabolites were classified into groups, ‘amino acid’, ‘organic acid’, ‘sugar’, ‘amine’ and ‘others’ which are colored with purple, blue, vermillion, neon yellow and green, respectively.

confidence interval varies from 0.5 to 12.3 cM with a mean of 3.9 cM (Table S3). A total of 279 genes with a selection signal were overlapped with those QTLs (Figure S6, Table S4). The percentage of metabolic variation explained by the QTLs ranged from 4.7% to 57.9% with a mean of 13.8%. Similarly, the phenotypic variation explained by each QTL in four tissues increased from early stage (YL) to late stage (MK) and the median R^2 values ranged from 9.2% in YL to 14.3% in MK (Figure 2f).

The influence of the different genetic backgrounds (Mo17 and teosinte) on primary metabolite abundance was examined. Most primary metabolite QTLs were characterized as displaying moderate additive effects (Figure 2f), which indicates that the accumulation of primary metabolites is probably influenced by many genes. The overwhelming proportion of the interactions between QTLs was positive in all four tissues. Furthermore, for QTLs with additive effects for increasing primary metabolite content, the teosinte alleles displayed greater effect than the Mo17 alleles. By contrast, the opposite was noted for kernel shape- and starch-related traits, that is the Mo17 alleles had greater effects than the teosinte alleles (Figure 3a). The regulation of metabolites and other traits measured in this study was clearly strongly specific at the spatiotemporal level (Figure 2a). Moreover, the uneven distribution of QTLs across the genome is even clearer when looking solely at the QTLs in a specific tissue or direction of additive effect (Figure 3b), with QTLs enriched in specific genome regions being likely to show the same directionality of effect.

Analysis of the observed QTL hotspots leads to clues concerning the process of maize domestication and improvement. The identified QTLs showed a significantly uneven distribution and two QTL hotspots (chromosomes 2 and 6, permutation test, $P < 0.01$) were identified (Figure 2b). These two hotspots are mainly composed of amino acid-related QTLs mapped in YK and MK tissues. The hotspot detected on chromosome 2 is mainly composed of amino acid-related QTLs in the YK tissue (Figure 2d). The hotspot on chromosome 6 encompassed QTLs for 19 metabolites (Figure 2e), 12 of which were amino acid-related in the MK. Gene Ontology (GO) analysis was conducted on the two hotspot intervals. Significantly enriched GO terms detected in the hotspot on chromosome 2 mainly concerned the negative regulation of biological process, cellulose metabolic process and glucan biosynthetic process. In this interval, some QTLs associated with maize starch content and starch viscosity were also mapped, but with opposite directionality from the primary metabolic QTLs (Figure 3a,b). Furthermore, 41 genes with domestication or improvement signals were also detected in this region, according to previous comparative population genomics analysis of large-scale re-sequencing of the genomes of maize and its wild relatives (Table S4)

(Hufford *et al.*, 2013). This may reflect that artificial selection for high yield forced changes in the regulation of primary metabolism, especially that of the amino acids. Within the other QTL hotspot on chromosome 6, GO terms with response to external stimulus were significantly enriched ($P < 0.001$). In previous studies, it was demonstrated that primary metabolites such as isoleucine, valine, threonine, 4-aminobutanoate, glycine, serine *myo*-inositol and tricarboxylic acid cycle intermediates increased following drought and/or salt stress (Henry *et al.*, 2015; Obata *et al.*, 2015). A QTL affecting kernel size with the favorable allele from Mo17 was also identified in this region (Figure 3b). We speculate that long-term domestication and breeding processes have reshaped the mechanism of amino acid regulation and metabolism to meet the increasing demand for high-yield maize.

A metabolic network comprising most of the primary metabolites detected in this study was adopted to demonstrate the additive effects and inter-metabolite correlations (Figure 4). This network includes 77% (50/65) of the identified primary metabolites, which can be roughly classified into three clusters. One cluster is composed of sugars including sucrose, glucose and fructose raffinose. The second, encompassing biosynthetic and catabolic pathways of amino acid metabolism, included 20 protein amino acids and two non-protein amino acids. The tricarboxylic acid cycle was the main component of the third cluster, with citrate, *cis*-aconitate and malate being involved. Within this network, the ratio of QTLs with favorable alleles coming from the male parent teosinte or female parent Mo17 was recorded and represented in the color bar adjacent to the metabolite name in red or blue, respectively (Figure 4). With the exception of the organic acids pyruvate and *cis*-aconitate, most metabolites had a greater number of QTLs with favorable alleles coming from the male parent teosinte. This suggests that there are many favorable alleles promoting primary metabolite content in the teosinte genome compared with the Mo17 genome – irrespective of tissue type and developmental stage. In a previous study, a maize ideotype with high grain yield was proposed that displayed low accumulation of soluble amino acids and carbohydrates in the leaves (Cañas *et al.*, 2017); intriguingly, the metabolome postulated for this ideotype is very similar to what we actually observed here for teosinte.

Even though genetic variance of quantitative traits was mainly characterized by additive gene effects (Hill *et al.*, 2008), epistatic effects have been detected and demonstrated to exert great influence on diverse characteristics in a wide range of crops. For example, tocopherol content in soybean (Liu *et al.*, 2017), fruit cracking in tomato (Capel *et al.*, 2017), barley yield (Xu and Jia, 2007), heading date in rice (Chen *et al.*, 2015) and primary metabolite abundance in maize (Wen *et al.*, 2015). In this research, epistatic effects that existed in every paired QTL for each primary

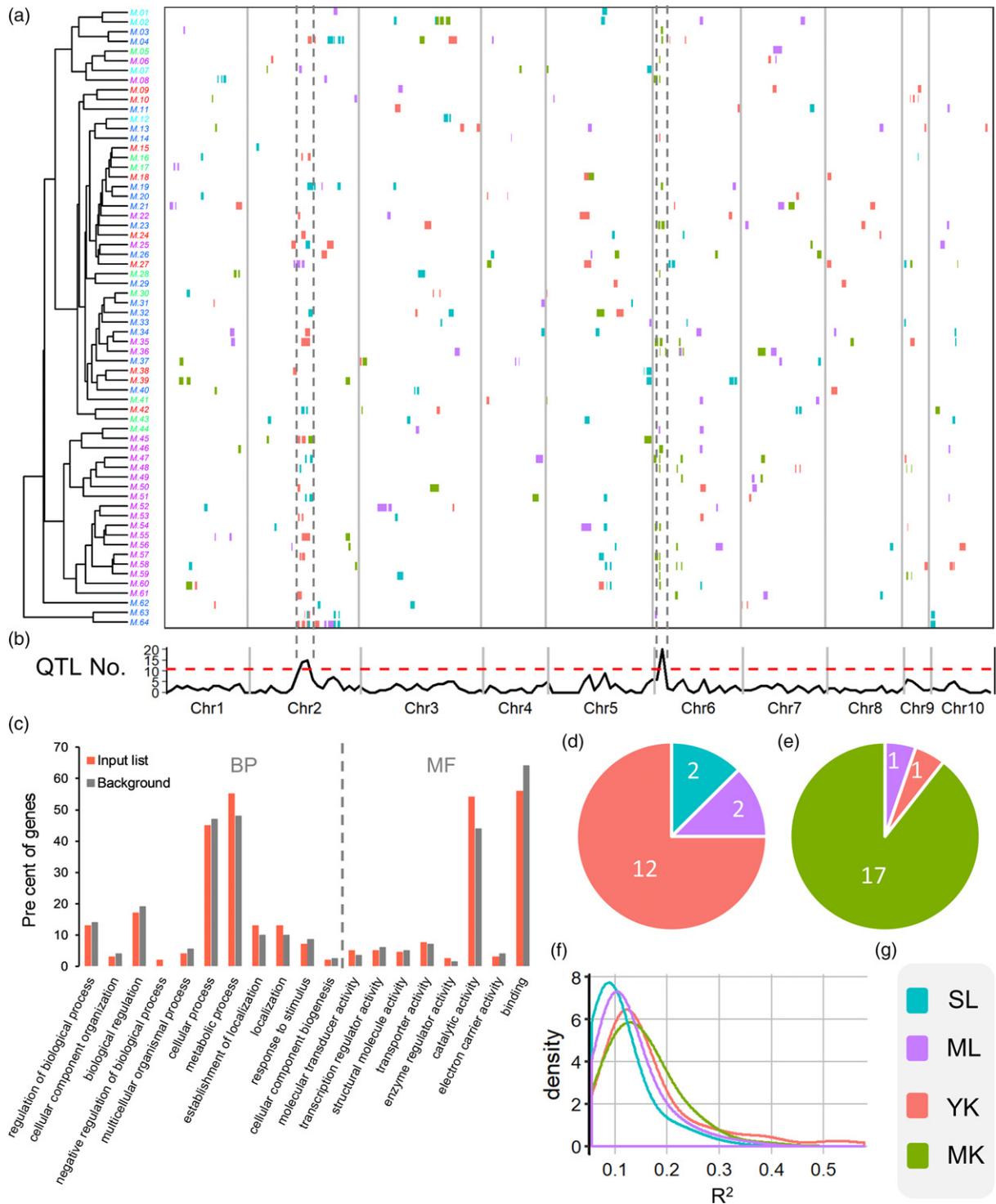


Figure 2. Summary of the locations of quantitative trait loci (QTLs) across the maize genome and their effects.

(a) Chromosomal distribution of primary metabolic QTLs identified in four tissues of the maize plant. The QTL confidence intervals are represented by rectangles, which are filled with four colors representing different tissues. Hierarchical clustering results were obtained based on primary metabolic values in four tissues. Corresponding metabolites are listed in Table S1. (b) Statistic of the number of QTLs across the genome. The window size is 4.3 Mb. (c) Enrichment analysis of Gene Ontology annotation of genes located in two hotspots with significant ($P < 0.05$) correlations between primary metabolic values and expression data in 10 lines. BP, biological process; MF, molecular function. (d) Pie plot showing the composition of primary metabolites with QTLs mapped in the hotspot on chromosome 2 classified according to tissues. (e) Pie plot showing composition of primary metabolites with QTLs mapped in the hotspot on chromosome 6 classified according to tissues. (f) Density plot of R^2 (explained phenotypic variation) values of QTLs identified in this study. (g) The legend used for (a), (d), (e) and (f).

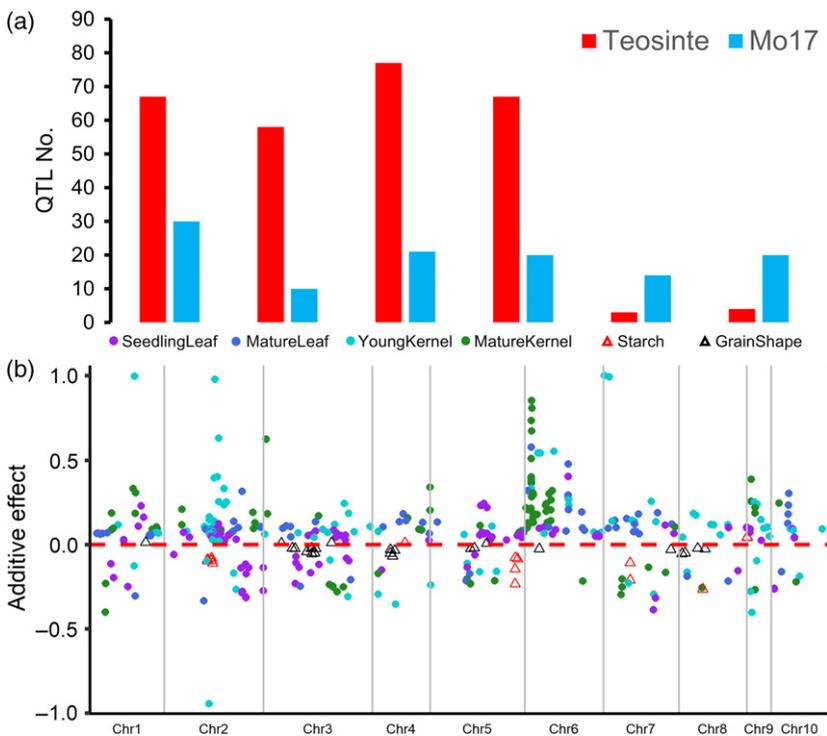


Figure 3. Summary of additive effect of the identified quantitative trait loci (QTLs).

(a) Bar plot showing the number of QTLs with positive and negative additive effects. Red bars represent the cases with a higher metabolite level in teosinte (i.e. with values greater than zero) and blue bars represent the cases with higher metabolic level in Mo17 (with values less than zero).

(b) Distribution of QTLs according to their physical location across the maize genome (x-axis). The y-axis represents the value of the additive effect of each QTL.

metabolite were evaluated (Figure 5a). A total of 13% (4.9–23% in four tissues) of primary metabolites were detected with significant ($P < 0.05$) epistatic interaction in different QTL regions (Figure 5b). Among them, chlorogenic acid (3-caffeoyl-*cis/trans* quinic acid)-related loci displayed epistatic interactions in two tissues, suggesting a conservative mechanism of metabolic regulation within those two tissues (Figure 5c). Furthermore, chlorogenic acid, quinic acid and caffeic acid are involved in reversible chemical reaction processes and displayed a higher epistatic effect (13.1%) in SL tissue and YK tissue than their average epistatic effect value. Interestingly, larger epistatic effects were detected in kernel tissues (about 12.5%) than in leaf tissues (about 4.9%), on average (Figure 5c).

Networks connecting primary metabolites, maize starch and grain shape traits

To further investigate the relationships between primary metabolites from each tissue and grain yield (starch content and grain shape), Pearson correlation coefficients (PCC) were obtained and presented in Table S5. In total, there were 299 (75, 49, 79 and 96 in SL, ML, YK and MK, respectively) significant ($P < 0.05$) correlations between primary metabolites and maize starch content or grain shape traits in the four tissues (Figure 6). In the seedling stage, almost two-thirds of the significant correlations were positive and focused on primary metabolites and grain-shape traits. However, at the ear leaf stage, more than 60% of the associations were negative and highly enriched between a

subset of primary metabolites (glyceric acid, nicotinic acid, xylitol, quinic acid, 3-phosphoglyceric acid) and grain shape and starch content. Chlorogenates (reversibly biosynthesized from quinic acid and shikimate) have been regarded as biomarkers for larger kernels in maize (Cañas *et al.*, 2017). In the present study, a negative correlation was also observed between quinic acid and grain starch content. Interestingly, 76% (28/37) of the pairwise primary metabolite–maize starch viscosity correlations were negative, while 83% (35/42) of the pairwise primary metabolite–kernel shape trait correlations were positive. Eight primary metabolites (2-oxoglutaric acid, arginine, glutamic acid, glutamine, isocitric acid, leucine, 3-caffeoyl-*cis* quinic acid and succinic acid) mainly involved in the shikimate pathway and the biosynthesis of amino acids derived from glutamate exhibited a completely opposite correlation from maize kernel starch viscosity and shape traits (Table S5). Furthermore, amino acids derived from glutamate have also been described as a biological indicator for selection for high grain yield in maize (Cañas *et al.*, 2017). In the MK, 44% (22/50) of the metabolites display positive correlations with maize starch viscosity characters and kernel shape traits. To our surprise, another 44% (22/50) of the metabolites have a negative correlation with the kernel test weight, which indicates that primary metabolite content may influence maize kernel density. These correlations were also confirmed by the co-location of QTLs, as shown in Figure 3(b). However, it is important to note that we cannot, at present, rule out that correlations between grain

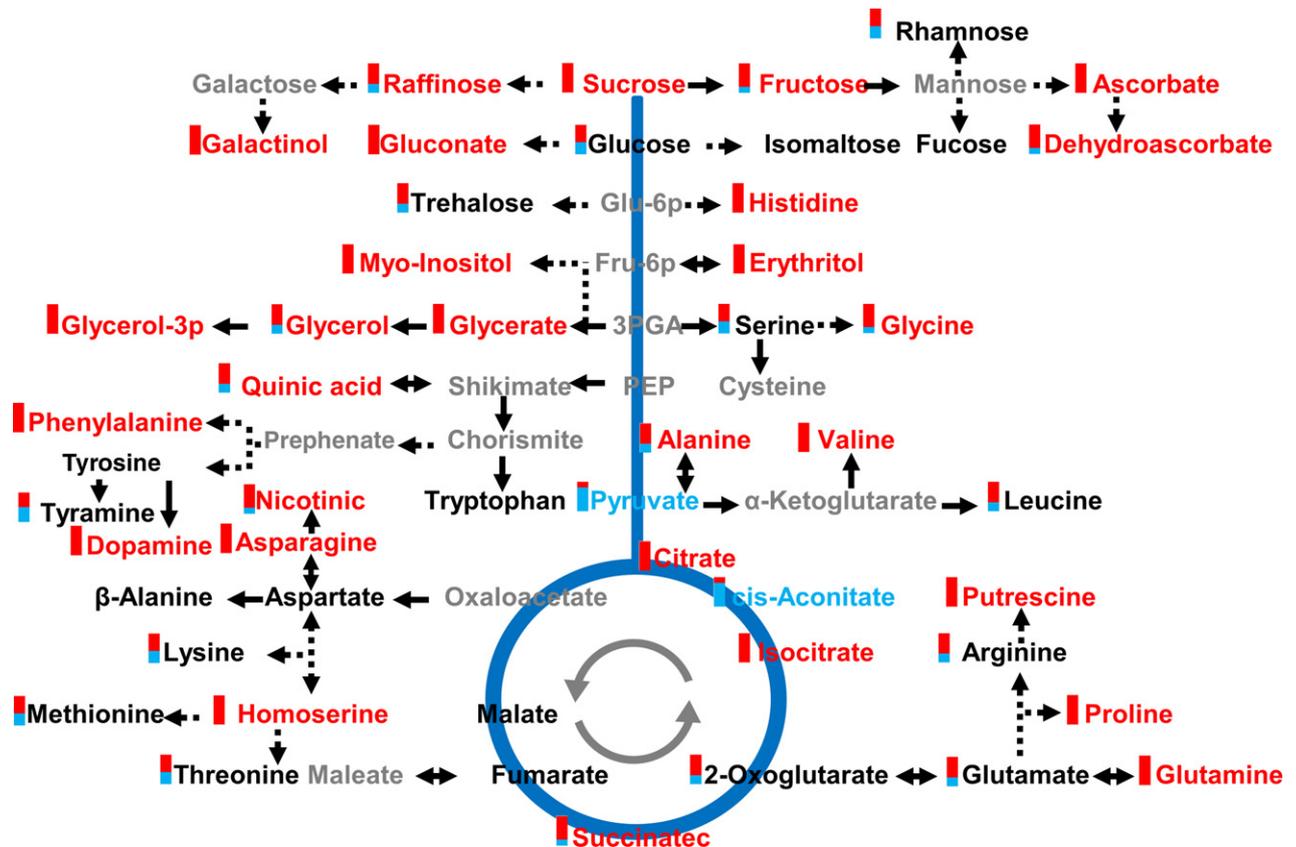


Figure 4. The teosinte–maize variation of the primary metabolic network.

For the primary metabolites shown in red, the majority of the identified quantitative trait loci (QTLs) have an additive effect with a value greater than zero (at least two-fold change between the number of QTLs having an additive effect with value greater than zero and the number of QTLs having an additive effect with value less than zero). For the primary metabolites shown in blue, the majority of the identified QTLs have an additive effect with value less than zero (at least two-fold change between the number of QTLs having an additive effect with value less than zero and the number of QTLs having an additive effect with value greater than zero). Metabolites shown in black did not have a significant difference between QTLs with additive effects greater than and less than zero (less than two-fold change). Metabolites that are not detected in this study are shown in grey. The color bar represents the proportion of these two kinds of QTL.

yield (starch content and grain shape traits) and primary metabolites are a consequence of pleiotropic effects or of linkage drag.

Combing multi-tissue transcriptome and genome-wide association studies to narrow the candidate genes

Genome-wide association studies (GWAS) are an efficient way to discover genes associated with phenotypes. To evaluate the accuracy of QTLs discovered by linkage mapping in the TM population and in order to narrow down the candidate genes, we compared these QTLs with the GWAS results of primary metabolites from a diverse association panel (Wen *et al.*, 2018). As would perhaps be expected, some degree of consistency was found between the QTL information we obtained in the TM population and the association panel. For example, both the tandem *ADT* genes (*GRMZM2G342895*, *GRMZM2G121546*) encoding arogenate dehydratase were cloned and their function validated in a previous study (Wen *et al.*, 2018). In the current

study, we also detected a significant QTL at the same locus in the TM population (Figure S7). A major QTL accounting for 31.4% of the variation in GABA content with the favorable allele from *mexicana* was mapped on chromosome 1 (Figure 7a). In the identical region, the *ZmGAD* gene (*GRMZM5G826838*) encoding glutamate decarboxylase was found by GWAS with six polymorphic markers, one located in the 5' untranslated (UTR) region and five located in the 3' UTR region (including the most significant one; Figure 7b,c). Also, a *cis*-QTL was identified for the expression level of *ZmGAD* (Figure 7d) and significant positive correlations were observed between the *ZmGAD* expression level and the GABA content in the SL ($r = 0.31$, $P = 1.3 \times 10^{-8}$) and MK ($r = 0.25$, $P = 4.5 \times 10^{-6}$) (Figure 7e). The glutamic acid could be decomposed into GABA under the function of glutamate decarboxylase (Figure 7f). These results imply that some primary metabolite QTL identified in this study may have already been used in modern maize breeding programs. Modulating gene

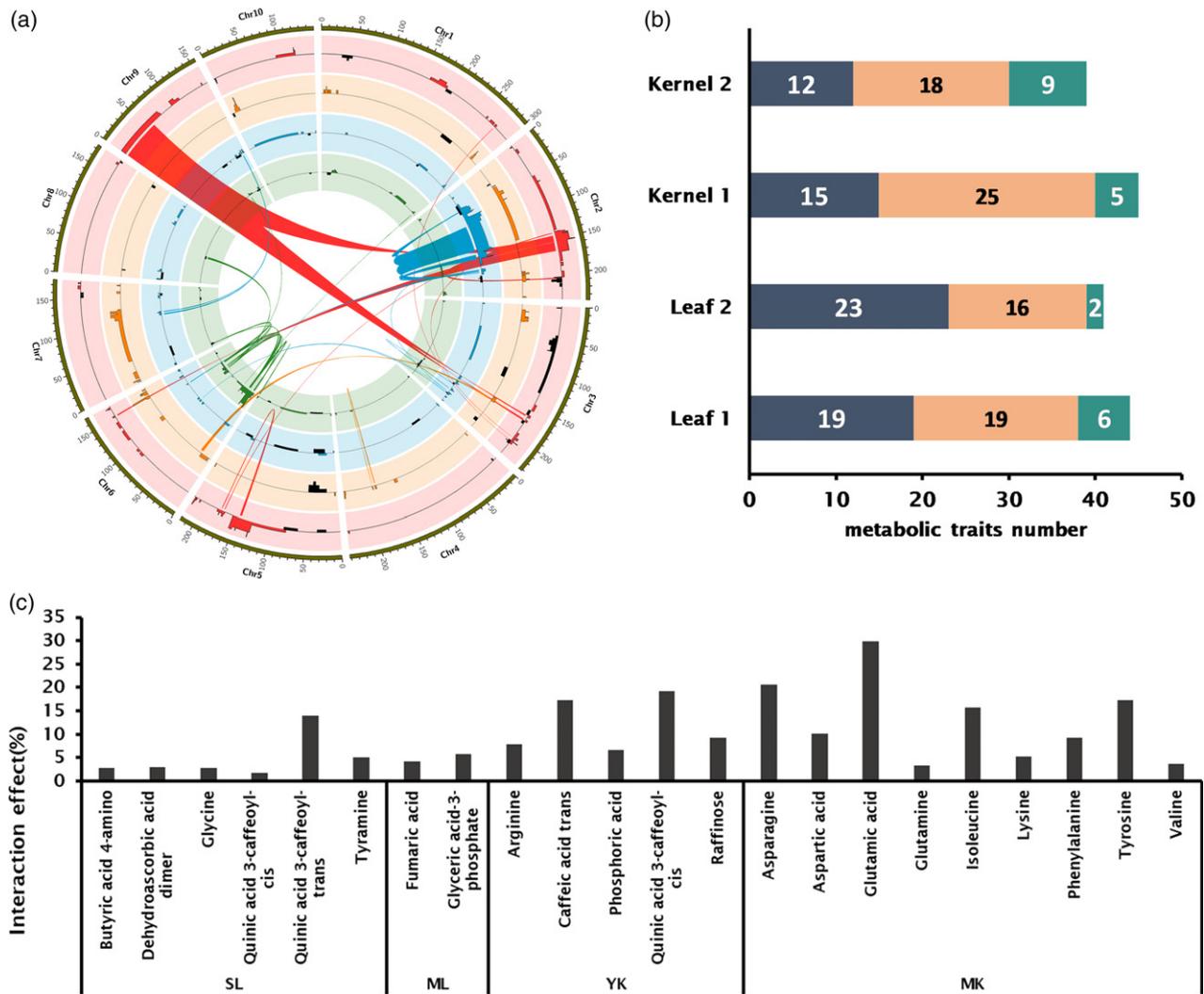


Figure 5. Summary of epistatic effects of primary metabolite quantitative trait loci (QTLs) identified in this study.

(a) Circos plot of the maize genome. The olive-colored track indicates the 10 chromosomes. Tick marks appear every 10 Mb on each chromosome. Four tracks from border to center (red histograms in a light red background, orange histograms in a light orange background, sky-blue histograms in a light sky-blue background and green histograms in a light green background) represent the number of QTLs detected at each site of the chromosomes for metabolites measured in seedling leaf, leaf at 15 DAP, young kernel and mature kernel, respectively. In each of the four tracks, outward histograms indicate a QTL having an additive effect with value greater than zero and inward histograms indicate a QTL having an additive effect with value less than zero.

(b) Statistics of metabolic traits with epistatic effects. Gray bars, number of metabolic traits with one QTL; orange bars, number of metabolic traits with epistatic effects detected; green bar, number of metabolic traits with no epistatic effects detected.

(c) Proportion of phenotypic variation explained by all epistatic interactions for each metabolite trait.

expression often represents an efficient mechanism for eliciting quantitative changes in gene function. In order to gain a better understanding of the regulation of primary metabolism, RNA sequencing experiments were conducted on three tissues (i.e. SL, YK and MK) in 10 lines selected from the TM population (Yang *et al.*, 2017). Correlations were subsequently calculated between primary metabolite abundance and gene expression. In total, 2726 transcript–metabolite correlations involving 63 primary metabolites and 2099 genes were identified at a permissive significance threshold of $P < 0.05$ (Table S4), with 434 of these 2099 genes displaying significantly different ($P < 0.05$)

expression levels between lines harboring Mo17 and teosinte alleles (Table S4). Co-expressed genes are potentially involved in the same biological process (Yu *et al.*, 2015). Thus, to investigate the biological function of these selected genes, *K*-means clustering was adopted in order to classify the genes on the basis of their expression patterns. As a result, nine clusters were identified, comprising between 87 and 634 genes (Figure S8 and Table S4). Subsequently, GO enrichment analyses were carried out on the genes of each cluster (Tian *et al.*, 2017). Significantly enriched GO terms were detected in seven clusters ($P < 0.05$). These terms were mainly involved in

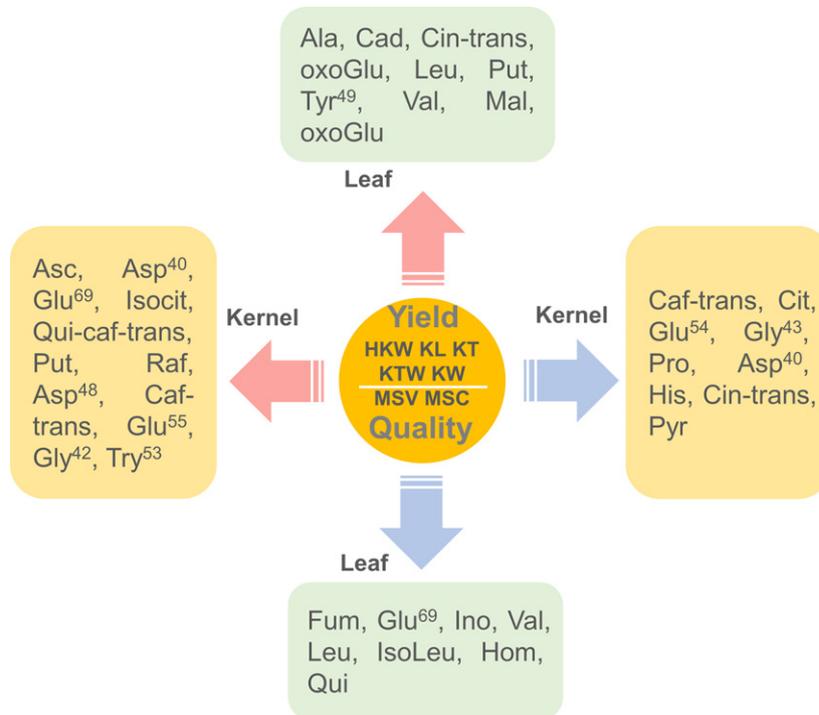


Figure 6. Simplified network of primary metabolites and yield-related traits.

Red and blue arrows represent the positive and negative correlations between yield-related traits and primary metabolic trait values, respectively. The light green backgrounds represent metabolites significantly correlated with yield-related traits in leaf tissues (red arrow, positive correlation; blue arrow, negative correlation) (i.e. seedling leaf and leaf at 15 days after pollination). Light yellow backgrounds represent metabolites significantly correlated with yield-related traits in kernel tissues (red arrow, positive correlation; blue arrow, negative correlation) (i.e. young kernel and mature kernel). A full list of correlation results is presented in Table S5.

Abbreviations: HKW, hundred kernel weight; KL, kernel length; KT, kernel thickness; KTW, kernel test weight; KW, kernel width; MSV, maize starch viscosity; MSC, maize starch content; Asc, ascorbic acid; Asp⁴⁰, asparagine; Glu⁶⁹, glucose; Isocit, isocitric acid; Put, putrescine; Qui-caf-trans, quinic acid 3-caffeoyl-*trans*; Raf, raffinose; Asp⁴⁸, aspartic acid; Caf-trans, caffeic acid-*trans*; Glu⁵⁵, glutamic acid; Gly⁴², glycerol; Try⁵³, tyramine; Cit, citric acid; Glu⁵⁴, glutamine; Gly⁴³, glyceric acid; Pro, proline; Cin-trans, cinnamic acid 4-hydroxy-*trans*; His, histidine; Pyr, pyruvic acid; Ala, alanine; Cad, cadaverine; OxoGlu, glutaric acid 2-oxo; Leu, leucine; Tyr⁴⁹, tyrosine; Val, valine; Mal, malic acid; Fum, fumaric acid; Ino, inositol myo; IsoLeu, isoleucine; Hom, homoserine; Qui, quinic acid. Superscripts represent the number of signals of the metabolites and detailed information can be found in Table S1.

carbohydrate metabolic process, biosynthetic process, protein metabolic process and phosphate metabolic process, but some molecular functions, including transferase activity, hydrolase activity and structural molecule activity, and some cellular components also appeared on the GO enrichment list, which can be found in Table S6. These data provide interesting hints to help us understand how this divergence arose and how it can be harnessed to enhance quality improvement of maize in the future.

DISCUSSION

Metabolomics has recently made several important contributions to our understanding of fundamental aspects of maize biology, including metabolic responses to climate change, pathogen attack, microorganism resistance and dynamic development of plants (Walker *et al.*, 2011; Marti *et al.*, 2013; Sun *et al.*, 2016; de Abreu e Lima *et al.*, 2018; Wen *et al.*, 2018). To date, comparisons of the metabolic differences between cultivated maize and teosinte have, however, been relatively limited (Flint-Garcia *et al.*, 2009).

In this study, large-scale primary metabolite detection was conducted in four different tissues of a maize–teosinte population alongside the collection of data on maize kernel shape traits and maize starch characteristics. This study revealed the dramatic differences in primary metabolite content and distribution between maize and teosinte (Figure 1). Hundreds of genome regions exhibited genetic contributions to the regulation of primary metabolite accumulation in both leaf and kernel tissues (Figure 2a). Importantly, the majority of the genetic contributions with an upregulation effect came from the *mexicana* alleles (Figure 3b). Previously, introgression of genome segments from *mexicana* into cultivated maize has been documented to result in large changes in composition of protein content (Wang *et al.*, 2008). However, this earlier study did not elucidate the genetic mechanism underlying this phenomenon. Also, teosinte appears to have greater resistance to a number of pests than its cultivated counterpart (Lange *et al.*, 2014). One previous study showed that teosinte alleles on chromosome 8 confer increased

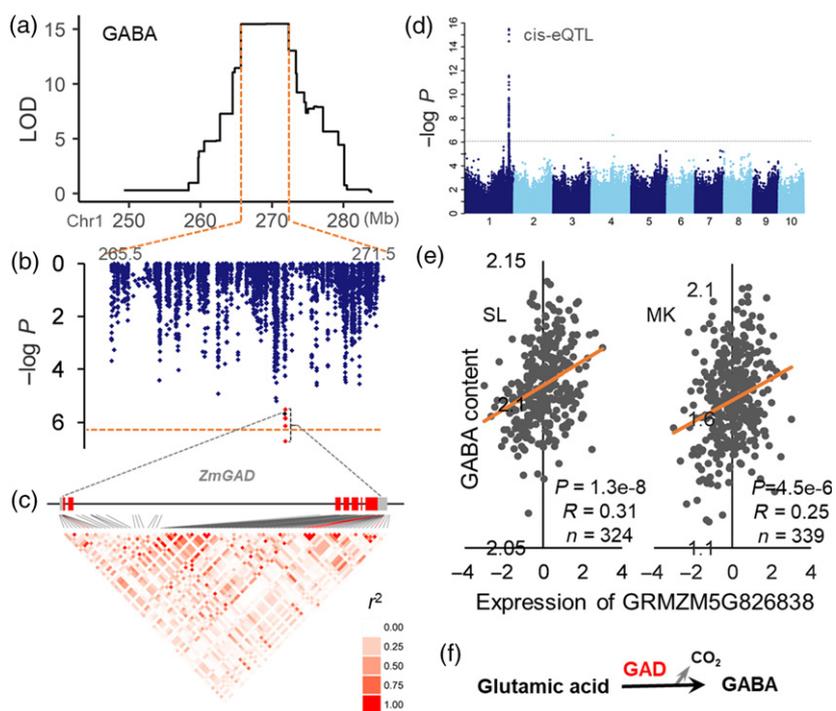


Figure 7. *ZmGAD* (*GRMZM5G826838*) as the candidate gene affecting 4-aminobutyric acid (GABA) content.

(a) Quantitative trait locus (QTL) mapping result of GABA content in the Mo17 and teosinte (*mexicana*) population (the TM population). (b) Scatterplot of association results between polymorphic markers in the GABA QTL confidence interval. Association analysis was performed using the mixed linear model controlling for the population structure (Q) and kinship (K). Six polymorphic markers locate in the 3' untranslated region (UTR) and 5' UTR region of *ZmGAD*, marked with black and red, respectively. (c) Structure of the *ZmGAD* gene and linkage disequilibrium plot showing the pair-wise r^2 value among polymorphic sites. Red lines show the position of six polymorphic markers mentioned above. (d) Manhattan plot of expression QTL analysis of *ZmGAD* expression. (e) Plot of Pearson correlation between the normalized expression level of the *ZmGAD* gene and the content of GABA in seedling leaf (SL) and mature kernel (MK). (f) Proposed GABA metabolic pathway in maize. GAD, glutamate decarboxylase.

resistance to gray leaf spot (Zhang *et al.*, 2017). Teosinte undergoes more stress than cultivated maize, and it is necessary for teosinte to have stronger resistance to biotic and abiotic stresses. These advantages are likely to be conveyed by proteins and secondary metabolites, which may benefit from the species' diverse genetic architecture (Lange *et al.*, 2014). While initially surprising, the observation that *mexicana* alleles have greater power to alter primary metabolite abundance, based on additive effect results from primary metabolite QTLs (Figure 4), is perhaps reasonable. Obviously, the process of maize domestication and improvement was accompanied with loss of genetic diversity (Xu *et al.*, 2017) and there are several possible reasons for this. (i) Genetic drift: modern maize is derived from landraces, which in turn were domesticated from teosinte (Shi and Lai, 2015). Due to limitations in the population size of teosinte and landraces, genetic drift could have occurred in every generation during these two processes. (ii) Selection: breeders focus more on the target traits they are interested in, such as plant morphological characteristics, grain yield and disease resistance. This leads to a genetic bottleneck wherein plants with no advantage in these traits are discarded and allelic diversity is lost across the genome. Furthermore, genes influencing target traits may be closely linked with genes nearby, for example primary metabolite QTLs and grain yield-related QTLs are co-localized (Figure 3b) so selection of target traits may influence the gene frequency in adjacent regions of the genome. (iii) Natural environmental stress: stressful environments can become a

driving force for plants to accumulate more metabolites as part of the adaptive response (Obata *et al.*, 2015). Cultivated maize has a much better growing environment, with humans controlling the entire growth cycle, than that to which its wild relatives were exposed. Therefore, there is much less impetus for cultivated maize to maintain a high allele frequency of genes involved in metabolite biosynthesis and accumulation. Results obtained in this study supported the consensus that wild relatives contain valuable germplasm for enhancing the genetic diversity reduced in the long selection process of maize domestication and improvement (Lange *et al.*, 2014; Yang *et al.*, 2017).

Genome research revealed that approximately 2–4% of maize genes have been influenced in the domestication process (Wright *et al.*, 2005; Hufford *et al.*, 2012), and several genes have completely transformed maize at the level of whole plant architecture, ear morphology and kernel structure (Shi and Lai, 2015). In this study, two hotspots were identified in which tens of primary metabolite QTLs were co-localized (Figure 2b), implying that the regulation of primary metabolites is influenced by several key genes. These two QTL hotspots provide good targets for further detailed study. It has previously been reported that domestication has significantly changed the expression levels of hundreds of gene as well as their co-expression networks (Swanson-Wagner *et al.*, 2012; Huang *et al.*, 2016). Furthermore, many of these differentially expressed genes are members of such co-expression networks and additionally annotated with biological functions in response to biotic

stimulus (Swanson-Wagner *et al.*, 2012). Intriguingly, several primary metabolite hotspots were also identified in a maize–maize (B73/BY804) population (Wen *et al.*, 2015). However, they did not overlap with the present two hotspots, which reflects the difference in network regulation between maize and its wild relatives. We believe that this is due to the evolution of differential regulation of the metabolic network, and particularly in amino acid metabolism, during domestication. In this vein, it is interesting to compare our findings with those of two recent studies which evaluated the domestication of the metabolome in wheat (Beleggia *et al.*, 2016) and tomato (Zhu *et al.*, 2018). Interestingly, five amino acids were found to accumulate with the increase in fruit size during crop improvement in the case of tomato. The case of wheat, however, is perhaps more pertinent to our study here, given that it is a far more similar species to maize than tomato. Beleggia *et al.* (2016) revealed that the domestication of emmer (primary domestication) was marked by a reduction in unsaturated fatty acids while the domestication of durum wheat (secondary domestication) was, like that of maize, marked by a massive reduction in amino acid abundance.

By combining metabolome, transcriptome, grain yield and quality trait data, we were able to comprehensively dissect primary metabolism using a maize by wild relative (*mexicana*) segregating population. We saw the difference in the maize and teosinte genomes and identified the same regions to affect the metabolites and grain yield simultaneously but with opposite effects. Similar phenomena were also observed in other crops (Beleggia *et al.*, 2016; Tieman *et al.*, 2017). Many favorable genes affecting different primary metabolites were hidden in the genomes of the wild relatives of maize, suggesting that they may have been lost during the breeding of maize for high yield. With the investigation of 895 inbred lines, about 10% of the maize genome showed evidence of introgression from *mexicana*; however, for any given individual the introgression region was less than 1%, and no identified functions have yet been determined for these introgression regions (Yang *et al.*, 2017). These results suggest that some favorable genes hidden in maize wild relatives may already have been used in modern maize breeding. Thus, wild relatives are clearly a great genetic resource for improvement of maize production and our present findings provide clear clues for future maize breeding strategies that would allow us adopt a balanced approach with the aim of maintaining/enhancing both yield and quality simultaneously.

EXPERIMENTAL PROCEDURES

Plant material

Teosinte (*Zea mays* ssp. *mexicana*) was the non-recurrent parent involved in a Mo17–Teo (BC₂F₇, 191 lines) population (here, the TM population) construction with Mo17 acting as the

recurrent parent. Briefly, teosinte pollen was crossed onto the ear filament of an elite maize line Mo17. The progenies were backcrossed twice with Mo17; the resultant offspring were self-pollinated until the construction of the TM population was complete. Details of the construction of the population were described in a previous study (Yang *et al.*, 2017). The TM population together with the parent Mo17 were planted at Huazhong Agricultural University field experiment station (Wuhan, 109°51' E, 18°25' N) in 2015. Each line was planted at random in a single 3-m-row field plot. Two leaf tissue samples [i.e. the ninth leaf, 50 days after sowing; ear leaf at 15 days after pollination (DAP)] and YK tissue samples (i.e. kernel at 15 DAP) were harvested from three living plants in each line; they were pooled and mixed and then immediately snap-frozen in liquid nitrogen. For long-term storage, the samples were kept at –80°C before further processing. The MK tissue samples were harvested from the dry ears. For quality control, each sample was taken from three individuals per field plot.

Gas chromatography time-of-flight MS-based quantification of primary metabolites

Samples were homogenized with a ball mill instrument (MM400, Retsch, <https://www.retsch.com/>), pre-cooled with liquid nitrogen. They were then extracted as previously described (Roessner *et al.*, 2001; Lisec *et al.*, 2006). Briefly, 700 μ l of 100% methanol and 30 μ l of ribitol (0.2 mg ml⁻¹ stock in water) were added to 50 mg of homogenized powder. Samples were agitated in a shaker at 70°C for 15 min, then centrifuged at 20 817g for 10 min. The supernatant was then collected and mixed with 375 μ l of chloroform and 750 μ l of water. The mixture was then centrifuged for 15 min at 1699g. One hundred and fifty microliters of the upper polar phase was taken and concentrated to dryness in a vacuum concentrator. After derivatization, 1 μ l of each sample was injected into a gas chromatography time-of-flight MS system. Gas chromatography was conducted with a 30-m MDN-35 column. The temperatures of injection, transfer line and ion source were set at 230°C, 250°C and 250°C, respectively. A library derived from the Golm Metabolome Database (Kopka *et al.*, 2005) was used to evaluate chromatograms and mass spectra as well as determine metabolite levels. Relative quantification for each sample was determined using a unique selected ion intensity for each metabolite. All primary metabolite values in four tissues can be found in supplementary data sets (Table S7, seedling stage leaves; Table S8, ear leaves at 15 DAP; Table S9, kernels at 15 DAP; Table S10, MKs).

Evaluation of maize starch viscosity

Maize samples were milled and subsequently sifted through a 0.15-mm sieve. Milling was conducted under the same condition to avoid effects of the moisture content on the rapid viscosity analyzer (RVA; RVA-TecMaster, Perten Instruments, <https://www.perten.com/>) measurement. The experiment was carried out strictly following the recommended method specific for corn starch. In brief, distilled water (25 \pm 0.01 ml) was added to the milled maize flour sample (3 \pm 0.01 g) in an aluminum RVA canister. A paddle was placed in the canister and rotated at 960 r.p.m. for 10 sec to disperse the maize sample. Then a constant paddle rotation of 160 r.p.m. was used for the viscosity evaluation. The sequential temperature curve for a 13-min test was as follows: (i) incubate at 50°C for 1.0 min; (ii) increase to 95°C using a 3.7°C min⁻¹ increment; (iii) keep at 95°C for 2.5 min; (iv) cool down to 50°C with a 3.8°C min⁻¹ decrement; (v) hold at 50°C for 2 min. Starch viscosity characteristics were composed of five primary components (pasting temperature, peak viscosity, time to peak, minimum viscosity,

final viscosity) and two secondary components (breakdown, setback) (Table S11). Breakdown values were calculated by subtracting minimum paste viscosity from peak viscosity, while setback values were calculated by subtracting minimum paste viscosity from final viscosity.

Measurement of maize kernel traits

Kernel-related traits [i.e. hundred kernel weight, HKW; kernel length, KL; kernel thickness, KT; kernel test weight (KTW; weight of 250 ml of kernels); kernel width, KW] of the TM population were collected from three independent and diverse environments (Henan, Yunnan and Chongqing) with two replications (2011 and 2012). Each line was grown in a single 3-m row with a planting density of 45 000 plants ha⁻¹. At least three randomly selected plants in each line were used for collection of kernel-related traits. To obtain accurate phenotypic values, raw kernel-related phenotypic data were transformed into best linear unbiased prediction (BLUP) values based on the lme4 package in the R software environment (version 3.3.2), and BLUP values were used for subsequent analyses, including QTL mapping and statistical analysis (Table S11).

Determination of total starch content

Total starch (including D-glucose and maltodextrins) was analyzed using a standard assay procedure (cat. no. K-TSTA, Megazyme, <https://www.megazyme.com/>), in which starch was quantitatively converted to glucose by digestion with a thermostable amylase and amyloglucosidase (DMSO format, AOAC official method 996.11). In brief, maize starch samples (100 mg, weighed accurately) were dispersed using 0.2 ml of ethanol (80% v/v) and stirred on a vortex mixer. Then, 2 ml of DMSO was added to the sample tube and stirred on a vortex mixer. Samples were incubated in a vigorously boiling water bath for 5 min followed by the addition of 3 ml of α -amylase enzyme solution (α -amylase enzyme content 0.1 ml, 3000 U ml⁻¹ at pH 6.5 and 40°C) and incubated in a vigorously boiling water bath for 6 min. After incubation at 100°C, samples were incubated at 50°C with 0.1 ml of amyloglucosidase (3300 U ml⁻¹) for 30 min. The total sample volume was accurately adjusted to 100 ml with deionized water, then centrifuged at 1000g for 10 min. The aliquot (0.1 ml) was transferred into a test tube, 3 ml of glucose determination reagent (GOPOD) was added, and then incubated at 50°C for 20 min. The absorbance of the samples as well as the D-glucose control was read at 510 nm against the reagent blank:

$$\text{Starch, \%} = \frac{\Delta A}{W \times F} \times FV \times \frac{162}{180} \times \frac{100}{100 - \text{moisture content(\%)}}$$

where ΔA is absorbance read against the reagent blank, F is the absorbance for 1 μg of standard D-glucose. W is the weight in milligrams of the analyzed flour, FV is the final volume in milliliters and 162/180 is a statistical adjustment from free D-glucose to anhydro-D-glucose.

Bin map construction and QTL mapping

Illumina MaizeSNP50 arrays containing 56 110 single nucleotide polymorphisms (SNPs) were employed for genotyping all the individuals and their parents. An in-house Perl script (https://github.com/panqingchun/linkage_map) developed based on Carthagene software (De Givry *et al.*, 2005) running on the Linux environment was used for the construction of genetic linkage maps (Pan *et al.*, 2016). A total of 12 390 SNP markers were classified into 1282 unique bins and each bin was represented by a single marker. The composite interval mapping (CIM) method was used in QTL

Cartographer software version 2.5 (Wang *et al.*, 2012) for QTL mapping analysis with the walk speed set at 0.5 cM. The QTL confidence interval spanned the genomic regions corresponding to one LOD drop from the peak (Table S12). More detailed information about bin map construction and QTL mapping information can be found in our previous studies (Wen *et al.*, 2015; Yang *et al.*, 2017).

Detection of epistatic QTLs for primary metabolites

Pairwise additive by additive epistatic interactions were evaluated for all QTLs identified from each primary metabolite. Epistatic interactions were determined by Two-way ANOVA by considering pairwise markers as two factors to find whether those two loci have significant ($P < 0.05$) mutual function (Yu *et al.*, 1997). Epistatic effects were obtained by comparing the residual of total effects of single-locus effects and two-locus interaction effects with all single-locus effects derived from no mutual function model. Circos was used for visualization of the epistatic results (Krzywinski *et al.*, 2009).

Statistical analysis and visualization

Statistical analysis and data arrangement were mainly conducted using scripts written in R software. Pearson correlation coefficients together with P -values were calculated with the Hmisc package. Hierarchical cluster analyses were realized with the agglomeration method of ward.D2. For K -means clustering, determination of K was made via multiple attempts in order to gain the optimal classification. A permutation test was used to determine the threshold value for QTL hotspots. In brief, all QTLs were randomly assigned to the genome. Then, the QTL number of all slide windows (4.3 cM, average length of QTL confidence intervals) was recorded. After the 1000-permutation test, the value that was significant ($P < 0.01$) was 10 for 350 QTLs across the genome. This process was completed with script running in R version 3.3.2.

Gene Ontology enrichment analysis

The GO enrichment analyses were implemented by agriGO v.2.0 (Tian *et al.*, 2017) through the Singular Enrichment Analysis tools with the Fisher statistical test and Yekutieli multitest adjustment, with the significance level set at 0.05. The filtered working gene list together with the gene annotation of maize was downloaded from MaizeGDB (<ftp://ftp.gramene.org/pub/gramene/maize/sequence.org/release-5b/filteredset/>) (Lawrence *et al.*, 2005).

ACKNOWLEDGEMENTS

JY was supported by the National Key Research and Development Program of China (2016YFD010100303), the National Natural Science Foundation of China (31525017) and Huazhong Agricultural University Scientific and Technological Self-Innovation Foundation. SA and ARF was supported by the EU Horizon 2020 grant PlantaSyst, 836 and the German Federal Ministry of Research and Education grant Full Throttle (BMBF, grant 837031B0205A).

CONFLICT OF INTEREST

The authors declare that they have no conflict of competing interest.

AUTHOR CONTRIBUTIONS

JY, ARF and WW designed and supervised this study. JL (Jiansheng Li) and XY developed the population. KL, WW, SA, HG, LW (Luxi Wang), WZ, JL (Jie Liu), YL and XW

collected the data and performed the data analysis. QP constructed the linkage map. WL managed the field experiments. YB and LW (Lothar Willmitzer) provided technical support and advice. KL, WW, AF and JY wrote the manuscript with input from the other authors.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article.

Figure S1. Distribution of metabolite level in the teosinte/Mo17 population in the seedling leaf tissue of each line.

Figure S2. Distribution of metabolite level in the teosinte/Mo17 population in the ear leaf at 15 days after pollination of each line.

Figure S3. Distribution of metabolite level in the teosinte/Mo17 population in the YK at 15 days after pollination of each line.

Figure S4. Distribution of metabolite level in the teosinte/Mo17 population in the mature kernel of each line.

Figure S5. Hierarchical clustering analysis of primary metabolites.

Figure S6. Quantitative trait loci with selection signals detected in the confidence interval.

Figure S7. The *ZmGAD* gene (*GRMZM5G826838*), a candidate genes affecting 4-aminobutyric acid (GABA) content, is co-localized on chromosome 1 with linkage mapping and association analysis in seedling leaf tissue.

Figure S8. The expression pattern of candidate genes revealed by significant correlation between primary metabolites and gene expression values.

Table S1. List of primary metabolites detected in two leaf tissues and two kernel tissues.

Table S2. List of metabolite–metabolite Pearson correlations in four tissues.

Table S3. List of quantitative trait locus mapping results of primary metabolites in four tissues.

Table S4. List of candidate genes selected on the basis of transcriptome data.

Table S5. Results of Pearson correlational analyses of metabolite–kernel shape/starch-related traits.

Table S6. Results of Gene Ontology enrichment of 2009 candidate genes classified into nine clusters.

Table S7. Normalized metabolite values of seedling stage leaves in Mo17–teosinte population.

Table S8. Normalized metabolite values of ear leaves at 15 days after pollination in the Mo17–teosinte population.

Table S9. Normalized metabolite values of kernels at 15 days after pollination in the Mo17–teosinte population.

Table S10. Normalized metabolite values of mature kernels in the Mo17–teosinte population.

Table S11. List of normalized maize starch-related traits and best linear unbiased prediction (BLUP) results of grain shape traits from five mutually independent environmental conditions.

Table S12. List of candidate genes and the peak bin for each quantitative trait locus detected in four tissues.

REFERENCES

de Abreu e Lima, F., Li, K., Wen, W., Yan, J., Nikoloski, Z., Willmitzer, L. and Brotman, Y. (2018) Unraveling lipid metabolism in maize with time-resolved multi-omics data. *Plant J.* **93**, 1102–1115. <https://doi.org/10.1111/tj.13833>

Beleggia, R., Rau, D., Laidò, G. *et al.* (2016) Evolutionary metabolomics reveals domestication-associated changes in tetraploid wheat kernels. *Mol. Biol. Evol.* **33**, 1740–1753. <https://doi.org/10.1093/molbev/msw050>.

Bomblies, K. and Doebley, J.F. (2006) Pleiotropic effects of the duplicate maize *FLORICAULA/LEAFY* genes *zfl1* and *zfl2* on traits under selection during maize domestication. *Genetics*, **172**, 519–531. <https://doi.org/10.1534/genetics.105.048595>.

Briggs, W.H., McMullen, M.D., Gaut, B.S. and Doebley, J. (2007) Linkage mapping of domestication loci in a large maize-teosinte backcross resource. *Genetics*, **177**, 1915–1928. <https://doi.org/10.1534/genetics.107.076497>.

Cañas, R.A., Yesbergenova-Cuny, Z., Simons, M. *et al.* (2017) Exploiting the genetic diversity of maize using a combined metabolomic, enzyme activity profiling, and metabolic modeling approach to link leaf physiology to kernel yield. *Plant Cell*, **29**, 919–943. <https://doi.org/10.1105/tpc.16.00613>.

Capel, C., Yuste-Lisbona, F.J., López-Casado, G., Angosto, T., Cuartero, J., Lozano, R. and Capel, J. (2017) Multi-environment QTL mapping reveals genetic architecture of fruit cracking in a tomato RIL *Solanum lycopersicum* × *S. pimpinellifolium* population. *Theor. Appl. Genet.* **130**, 213–222. <https://doi.org/10.1007/s00122-016-2809-9>.

Chen, J., Li, X., Cheng, C., Wang, Y., Qin, M., Zhu, H., Zeng, R., Fu, X., Liu, Z. and Zhang, G. (2015) Characterization of epistatic interaction of qtls LH8 and EH3 controlling heading date in rice. *Sci. Rep.* **4**, 4263. <https://doi.org/10.1038/srep04263>.

De Givry, S., Bouchez, M., Chabrier, P., Milan, D. and Schiex, T. (2005) CARHTA GENE: multipopulation integrated genetic and radiation hybrid mapping. *Bioinformatics*, **21**, 1703–1704. <https://doi.org/10.1093/bioinformatics/bti222>.

Doebley, J., Stec, A. and Hubbard, L. (1997) The evolution of apical dominance in maize. *Nature*, **386**, 485–488. <https://doi.org/10.1038/386485a0>.

Duvick, D.N. (2005) Genetic progress in yield of United States maize (*Zea mays* L.). *Maydica*, **50**, 193–202. https://doi.org/10.1300/J064v25n04_08.

Fang, Z., Pyhäjärvi, T., Weber, A.L., Dawe, R.K., Glaubitz, J.C., González, J.J., Ross-Ibarra, C., Doebley, J., Morrell, P.L. and Ross-Ibarra, J. (2012) Megabase-scale inversion polymorphism in the wild ancestor of maize. *Genetics*, **191**, 883–894. <https://doi.org/10.1534/genetics.112.138578>.

Fernie, A.R. and Tohge, T. (2017) The genetics of plant metabolism. *Annu. Rev. Genet.* **51**, 287–310. <https://doi.org/10.1038/ng1815>.

Flint-Garcia, S.A., Bodnar, A.L. and Scott, M.P. (2009) Wide variability in kernel composition, seed characteristics, and zein profiles among diverse maize inbreds, landraces, and teosinte. *Theor. Appl. Genet.* **119**, 1129–1142. <https://doi.org/10.1007/s00122-009-1115-1>.

Gallavotti, A., Zhao, Q., Kyozuka, J., Meeley, R.B., Ritter, M.K., Doebley, J.F., Pè, M.E. and Schmidt, R.J. (2004) The role of *barren stalk1* in the architecture of maize. *Nature*, **432**, 630–635. <https://doi.org/10.1038/nature03148>.

Guo, L., Wang, X., Zhao, M. *et al.* (2018) Stepwise cis-regulatory changes in ZCN8 contribute to maize flowering-time adaptation. *Curr. Biol.* **28**, 3005–3015. <https://doi.org/10.1016/j.cub.2018.07.029>.

Henry, C., Bledsoe, S.W., Griffiths, C.A., Kollman, A., Paul, M.J., Sakr, S. and Lagrimini, L.M. (2015) Differential role for trehalose metabolism in salt-stressed maize. *Plant Physiol.* **169**, 1072–1089. <https://doi.org/10.1104/pp.15.00729>.

Hill, W.G., Goddard, M.E. and Visscher, P.M. (2008) Data and theory point to mainly additive genetic variance for complex traits. *PLoS Genet.* **4**, e1000008. <https://doi.org/10.1371/journal.pgen.1000008>.

Huang, J., Gao, Y., Jia, H. and Zhang, Z. (2016) Characterization of the teosinte transcriptome reveals adaptive sequence divergence during maize domestication. *Mol. Ecol. Resour.* **16**, 1465–1477. <https://doi.org/10.1111/1755-0998.12526>.

Hufford, M.B., Xu, X., Van Heerwaarden, J. *et al.* (2012) Comparative population genomics of maize domestication and improvement. *Nat. Genet.* **44**, 808–811. <https://doi.org/10.1038/ng.2309>.

Hufford, M.B., Lubinsky, P., Pyhäjärvi, T., Devengeno, M.T., Ellstrand, N.C. and Ross-Ibarra, J. (2013) The genomic signature of crop-wild introgression in maize. *PLoS Genet.* **9**, e1003477. <https://doi.org/10.1371/journal.pgen.1003477>.

Kopka, J., Schauer, N., Krueger, S. *et al.* (2005) GMD@CSB.DB: the Golm metabolome database. *Bioinformatics*, **21**, 1635–1638. <https://doi.org/10.1093/bioinformatics/bti236>.

Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., Jones, S.J. and Marra, M.A. (2009) Circo: an information aesthetic for comparative genomics. *Genome Res.* **19**, 1639–1645. <https://doi.org/10.1101/gr.092759.109>.

- Lange, E.S.D., Balmer, D., Mauch-Mani, B. and Turlings, T.C.J. (2014) Insect and pathogen attack and resistance in maize and its wild ancestors, the teosintes. *New Phytol.* **204**, 329–341. <https://doi.org/10.1111/nph.13005>.
- Lawrence, C.J., Seigfried, T.E. and Brendel, V. (2005) The maize genetics and genomics database. The community resource for access to diverse maize data. *Plant Physiol.* **138**, 55–58. <https://doi.org/10.1104/pp.104.059196>
- Lennon, J.R., Krakowsky, M., Goodman, M., Flint-Garcia, S. and Balint-Kurti, P.J. (2016) Identification of alleles conferring resistance to gray leaf spot in maize derived from its wild progenitor species teosinte. *Crop Sci.* **56**, 209–218. <https://doi.org/10.2135/cropsci2014.07.0468>.
- Lennon, J.R., Krakowsky, M., Goodman, M., Flint-Garcia, S. and Balint-Kurti, P.J. (2017) Identification of teosinte alleles for resistance to southern leaf blight in near isogenic maize lines. *Crop Sci.* **57**, 1973–1983. <https://doi.org/10.2135/cropsci2016.12.0979>.
- Lisec, J., Schauer, N., Kopka, J., Willmitzer, L. and Fernie, A.R. (2006) Gas chromatography mass spectrometry-based metabolite profiling in plants. *Nat. Protoc.* **1**, 387–396. <https://doi.org/10.1038/nprot.2006.59>.
- Liu, H., Wang, X., Warburton, M.L. et al. (2015) Genomic, transcriptomic, and phenomic variation Reveals the complex adaptation of modern maize breeding. *Mol. Plant*, **8**, 871–884. <https://doi.org/10.1016/j.molp.2015.01.016>.
- Liu, Z., Cook, J., Melia-Hancock, S. et al. (2016) Expanding maize genetic resources with predomestication alleles: maize-teosinte introgression populations. *Plant Genome*, **9**, 1–11. <https://doi.org/10.3835/plantgenome2015.07.0053>.
- Liu, H., Cao, G., Han, Y., Jiang, Z., Zhao, H. and Li, W. (2017) Identification of the QTL underlying the vitamin E content of soybean seeds. *Plant Breed.* **136**, 147–154. <https://doi.org/10.1111/pbr.12454>.
- Lu, X., Yang, L., Yu, M., Lai, J., Wang, C., McNeil, D., Zhou, M. and Yang, C. (2017) A novel *Zea mays* ssp. *mexicana* L. MYC-type ICE-like transcription factor gene *ZmmICE1*, enhances freezing tolerance in transgenic *Arabidopsis thaliana*. *Plant Physiol. Biochem.* **113**, 78–88. <https://doi.org/10.1016/j.plaphy.2017.02.002>.
- Luo, J. (2015) Metabolite-based genome-wide association studies in plants. *Curr. Opin. Plant Biol.* **24**, 31–38. <https://doi.org/10.1016/j.pbi.2015.01.006>.
- Maag, D., Erb, M., Bernal, J.S., Wolfender, J.L., Turlings, T.C.J. and Glauser, G. (2015) Maize domestication and anti-herbivore defences: leaf-specific dynamics during early ontogeny of maize and its wild ancestors. *PLoS ONE*, **10**, e0135722. <https://doi.org/10.1371/journal.pone.0135722>.
- Marti, G., Erb, M., Boccard, J., Glauser, G., Doyen, G.R., Villard, N., Robert, C.A.M., Turlings, T.C.J., Rudaz, S. and Wolfender, J.L. (2013) Metabolomics reveals herbivore-induced metabolites of resistance and susceptibility in maize leaves and roots. *Plant, Cell Environ.* **36**, 621–639. <https://doi.org/10.1111/pce.12002>.
- Matsuoka, Y., Vigouroux, Y., Goodman, M.M., Sanchez, G.J., Buckler, E. and Doebley, J. (2002) A single domestication for maize shown by multi-locus microsatellite genotyping. *Proc. Natl Acad. Sci. USA*, **99**, 6080–6084. <https://doi.org/10.1073/pnas.052125199>.
- Obata, T., Witt, S., Lisec, J., Palacios-Rojas, N., Florez-Sarasa, I., Araus, J.L., Cairns, J.E., Yousfi, S. and Fernie, A.R. (2015) Metabolite profiles of maize leaves in drought, heat and combined stress field trials reveal the relationship between metabolism and grain yield. *Plant Physiol.* **169**, 2665–2683. <https://doi.org/10.1104/pp.15.01164>.
- Pan, Q., Li, L., Yang, X., Tong, H., Xu, S., Li, Z., Li, W., Muehlbauer, G.J., Li, J. and Yan, J. (2016) Genome-wide recombination dynamics are associated with phenotypic variation in maize. *New Phytol.* **210**, 1083–1094. <https://doi.org/10.1111/nph.13810>.
- Piperno, D.R., Ranere, A.J., Holst, I., Iriarte, J. and Dickau, R. (2009) Starch grain and phytolith evidence for early ninth millennium B.P. maize from the Central Balsas River Valley, Mexico. *Proc. Natl Acad. Sci. USA*, **106**, 5019–5024. <https://doi.org/10.1073/pnas.0812525106>.
- Roessner, U., Luedemann, A., Brust, D., Fiehn, O., Linke, T., Willmitzer, L. and Fernie, A.R. (2001) Metabolic profiling allows comprehensive phenotyping of genetically or environmentally modified plant systems. *Plant Cell*, **13**, 11–29. <https://doi.org/10.2307/3871150>.
- Schauer, N., Semel, Y., Roessner, U. et al. (2006) Comprehensive metabolic profiling and phenotyping of interspecific introgression lines for tomato improvement. *Nat. Biotechnol.* **24**, 447–454. <https://doi.org/10.1038/nbt1192>.
- Shi, J. and Lai, J. (2015) Patterns of genomic changes with crop domestication and breeding. *Curr. Opin. Plant Biol.* **24**, 47–53. <https://doi.org/10.1016/j.pbi.2015.01.008>.
- Sun, C.X., Gao, X.X., Li, M.Q., Fu, J.Q. and Zhang, Y.L. (2016) Plastic responses in the metabolome and functional traits of maize plants to temperature variations. *Plant Biol.* **18**, 249–261. <https://doi.org/10.1111/plb.12378>.
- Swanson-Wagner, R., Briskine, R., Schaefer, R., Hufford, M.B., Ross-Ibarra, J., Myers, C.L., Tiffin, P. and Springer, N.M. (2012) Reshaping of the maize transcriptome by domestication. *Proc. Natl Acad. Sci. USA*, **109**, 11878–11883. <https://doi.org/10.1073/pnas.1201961109>.
- Takahashi, C.G., Kalns, L.L. and Bernal, J.S. (2012) Plant defense against fall armyworm in micro-sympatric maize (*Zea mays* ssp. *mays*) and balsas teosinte (*Zea mays* ssp. *parviglumis*). *Entomol. Exp. Appl.* **145**, 191–200. <https://doi.org/10.1111/eea.12004>.
- Tian, T., Liu, Y., Yan, H., You, Q., Yi, X., Du, Z., Xu, W. and Su, Z. (2017) AgriGO v2.0: a GO analysis toolkit for the agricultural community, 2017 update. *Nucleic Acids Res.* **45**, W122–W129. <https://doi.org/10.1093/nar/gkx382>.
- Tieman, D., Zhu, G., Resende, M.F.R. et al. (2017) A chemical genetic roadmap to improved tomato flavor. *Science*, **355**, 391–394. <https://doi.org/10.1126/science.aal1556>.
- Venkatesh, T.V., Chassy, A.W., Fiehn, O., Flint-Garcia, S., Zeng, Q., Skogerston, K. and Harrigan, G.G. (2016) Metabolomic assessment of key maize resources: GC-MS and NMR profiling of grain from B73 hybrids of the Nested Association Mapping (NAM) founders and of geographically diverse landraces. *J. Agric. Food Chem.* **64**, 2162–2172. <https://doi.org/10.1021/acs.jafc.5b04901>.
- Verlaan, M.G., Hutton, S.F., Ibrahim, R.M., Kormelink, R., Visser, R.G., Scott, J.W., Edwards, J.D. and Bai, Y. (2013) The tomato yellow leaf curl virus resistance genes *ty-1* and *ty-3* are allelic and code for DFDGD-class RNA-dependent RNA polymerases. *PLoS Genet.* **9**, e1003399. <https://doi.org/10.1371/journal.pgen.1003399>.
- Vollbrecht, E., Springer, P.S., Goh, L., Buckler, E.S. and Martienssen, R. (2005) Architecture of floral branch systems in maize and related grasses. *Nature*, **436**, 1119–1126. <https://doi.org/10.1038/nature03892>.
- Walker, V., Bertrand, C., Bellvert, F., Moëgne-Loccoz, Y., Bally, R. and Comte, G. (2011) Host plant secondary metabolite profiling shows a complex, strain dependent response of maize to plant growth promoting rhizobacteria of the genus *Azospirillum*. *New Phytol.* **189**, 494–506. <https://doi.org/10.1111/j.1469-8137.2010.03484.x>.
- Wang, H., Nussbaum-Wagler, T., Li, B., Zhao, Q., Vigouroux, Y., Faller, M., Bombliès, K., Lukens, L. and Doebley, J.F. (2005) The origin of the naked grains of maize. *Nature*, **436**, 714–719. <https://doi.org/10.1038/nature03863>.
- Wang, L., Xu, C., Qu, M. and Zhang, J. (2008) Kernel amino acid composition and protein content of introgression lines from *Zea mays* ssp. *mexicana* into cultivated maize. *J. Cereal Sci.* **48**, 387–393. <https://doi.org/10.1016/j.jcs.2007.09.014>.
- Wang, S., Basten, C.J. and Zeng, Z.B. (2012) *Windows QTL Cartographer 2.5*. Raleigh, NC: Department of Statistics, North Carolina State University. <http://statgen.ncsu.edu/qtlcart/winqtlcart.htm>.
- Wen, W., Li, D., Li, X. et al. (2014) Metabolome-based genome-wide association study of maize kernel leads to novel biochemical insights. *Nat. Commun.* **5**, 3438. <https://doi.org/10.1038/ncomms4438>.
- Wen, W., Li, K., Alseekh, S. et al. (2015) Genetic determinants of the network of primary metabolism and their relationships to plant performance in a maize recombinant inbred line population. *Plant Cell*, **27**, 1839–1856. <https://doi.org/10.1105/tpc.15.00208>.
- Wen, W., Jin, M., Li, K. et al. (2018) An integrated multi-layered analysis of the metabolic networks of different tissues uncovers key genetic components of primary metabolism in maize. *Plant J.* **93**, 1116–1128. <https://doi.org/10.1111/tpj.13835>.
- Wright, S.I., Bi, I.V., Schroeder, S.G., Yamasaki, M., Doebley, J.F., McMullen, M.D. and Gaut, B.S. (2005) The effects of artificial selection on the maize genome. *Science*, **308**, 1310–1314. <https://doi.org/10.1126/science.1107891>.
- Xu, S. and Jia, Z. (2007) Genome-wide analysis of epistatic effects for quantitative traits in barley. *Genetics*, **175**, 1955–1963. <https://doi.org/10.1534/genetics.106.066571>.
- Xu, G., Wang, X., Huang, C. et al. (2017) Complex genetic architecture underlies maize tassel domestication. *New Phytol.* **214**, 852–864. <https://doi.org/10.1111/nph.14400>.

- Yang, N., Xu, X.W., Wang, R.R. et al.** (2017) Contributions of *Zea mays* subspecies mexicana haplotypes to modern maize. *Nat. Commun.* **8**, 1874. <https://doi.org/10.1038/s41467-017-02063-5>.
- Yu, S.B., Li, J.X., Xu, C.G., Tan, Y.F., Gao, Y.J., Li, X.H., Zhang, Q. and Maroof, M.A.S.** (1997) Importance of epistasis as the genetic basis of heterosis in an elite rice hybrid. *Proc. Natl Acad. Sci. USA*, **94**, 9226–9231. <https://doi.org/10.1073/pnas.94.17.9226>.
- Yu, C.P., Chen, S.C.C., Chang, Y.M. et al.** (2015) Transcriptome dynamics of developing maize leaves and genomewide prediction of cis elements and their cognate transcription factors. *Proc. Natl Acad. Sci. USA*, **112**, E2477–E2486. <https://doi.org/10.1073/pnas.1500605112>.
- Zamir, D.** (2001) Improving plant breeding with exotic genetic libraries. *Nat. Rev. Genet.* **2**, 983–989. <https://doi.org/10.1038/35103589>.
- Zavala-López, M., López-Tavera, E., Figueroa-Cárdenas, J.D.D., Serna-Saldivar, S.O. and García-Lara, S.** (2018) Screening of major phenolics and antioxidant activities in teosinte populations and modern maize types. *J. Cereal Sci.* **79**, 276–285. <https://doi.org/10.1016/j.jcs.2017.11.007>.
- Zhang, N., Gibon, Y., Wallace, J.G. et al.** (2015) Genome-wide association of carbon and nitrogen metabolism in the maize nested association mapping population. *Plant Physiol.* **168**, 575–583. <https://doi.org/10.1104/pp.15.00025>.
- Zhang, X., Yang, Q., Rucker, E., Thomason, W. and Balint-Kurti, P.** (2017) Fine mapping of a quantitative resistance gene for gray leaf spot of maize (*Zea mays* L.) derived from teosinte (*Z. mays* ssp. *parviglumis*). *Theor. Appl. Genet.* **130**, 1285–1295. <https://doi.org/10.1007/s00122-017-2888-2>.
- Zhu, G., Wang, S., Huang, Z. et al.** (2018) Rewiring of the fruit metabolome in tomato breeding. *Cell*, **172**, 249–261. <https://doi.org/10.1016/j.cell.2017.12.019>.