

An integrated multi-layered analysis of the metabolic networks of different tissues uncovers key genetic components of primary metabolism in maize

Weiwei Wen^{1,2,3}, Min Jin¹, Kun Li¹, Haijun Liu¹, Yingjie Xiao¹, Mingchao Zhao^{1,4}, Saleh Alseekh², Wenqiang Li¹, Francisco de Abreu e Lima², Yariv Brotman^{2,5}, Lothar Willmitzer², Alisdair R. Fernie^{2,*} and Jianbing Yan^{1,*}

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

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

³Key Laboratory of Horticultural Plant Biology (Ministry of Education), Huazhong Agricultural University, Wuhan 430070, China,

⁴National Center of Plant Gene Research, Huazhong Agricultural University, Wuhan 430070, China, and

⁵Department of Life Sciences, Ben-Gurion University of the Negev, Beersheba, Israel

Received 15 May 2017; revised 20 December 2017; accepted 8 January 2018; published online 30 January 2018.

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

Weiwei Wen and Min Jin are the authors contributed equally.

SUMMARY

Primary metabolism plays a pivotal role in normal plant growth, development and reproduction. As maize is a major crop worldwide, the primary metabolites produced by maize plants are of immense importance from both calorific and nutritional perspectives. Here a genome-wide association study (GWAS) of 61 primary metabolites using a maize association panel containing 513 inbred lines identified 153 significant loci associated with the level of these metabolites in four independent tissues. The genome-wide expression level of 760 genes was also linked with metabolite levels within the same tissue. On average, the genetic variants at each locus or transcriptional variance of each gene identified here were estimated to have a minor effect (4.4–7.8%) on primary metabolic variation. Thirty-six loci or genes were prioritized as being worthy of future investigation, either with regard to functional characterization or for their utility for genetic improvement. This target list includes the well-known opaque 2 (*O2*) and *lkr/sdh* genes as well as many less well-characterized genes. During our investigation of these 36 loci, we analyzed the genetic components and variations underlying the trehalose, aspartate and aromatic amino acid pathways, thereby functionally characterizing four genes involved in primary metabolism in maize.

Keywords: *Zea mays*, primary metabolism, amino acids, trehalose, GWAS, functional verification.

INTRODUCTION

Maize is a major crop for food, feed and biofuel supply, and has the highest global grain production of all crops worldwide (Haley, 2011). Furthermore, it has recently been designated as a representative model C₄ plant (Pick *et al.*, 2011; Wang *et al.*, 2014). Due to its agricultural and economic value, improvement of either the protein or oil content in maize has been the focus of intensive grain quality breeding programs.

Recent advances in metabolite profiling technologies as well as genetic and systems biology-based approaches have extended the breeding portfolio beyond the traditional improvement targets of oil and protein to a wide variety of chemical compounds, including essential amino acids, vitamins, antioxidants and other metabolites of

physiological and nutritional importance (Wen *et al.*, 2016a). The concept has arisen of using metabolome-assisted techniques to bridge the gap between genotype and complex traits such as yield and biomass production (Fernie and Schauer, 2009; Fukushima and Kusano, 2013; Wen *et al.*, 2015; Cañas *et al.*, 2017). Owing to these advances, studies of plant metabolism, such as the comparative study of individuals within large populations, have increased in a manner which is likely to facilitate future breeding of both high-yielding and nutritionally rich crops (Keurentjes, 2009; Saito and Matsuda, 2010; Chan *et al.*, 2011; Wen *et al.*, 2014). However, detailed systematic investigations of the natural metabolic diversity of maize and its underlying genetic basis are still required.

Plant primary metabolism profoundly influences plant growth, development and reproduction (Schauer *et al.*, 2006, 2008). Metabolites such as carbohydrates, amino acids and organic acids accumulate in sink organs (e.g. seeds, fruits, and tubers) underlying a wide range of crop quality traits (Alonso-Blanco *et al.*, 2009). Deciphering the influence of genetics on primary metabolism is thus essential to strategies focused on metabolic engineering of chemical composition or biomass production. Most agriculturally and economically important traits of maize, including primary metabolic traits, are determined by multiple quantitative trait loci (QTLs; Wen *et al.*, 2015, 2016a). Linkage mapping and genome-wide association study (GWAS) are commonly used to dissect complex traits and precisely locate and characterize these functional loci, paving the way for crop improvement via marker-assisted selection or biotechnology-aided breeding. Taking advantage of linkage mapping using a recombinant inbred line (RIL) population, we have previously identified a large number of QTLs associated with the variation of 79 primary metabolites which enriched our understanding of the genetic basis underlying primary metabolic variation in multiple maize tissues (Wen *et al.*, 2015). GWAS has been used in plants over the last decade, with screens covering a large variety of phenotypic traits including large-scale primary metabolic traits (Riedelsheimer *et al.*, 2012; Wen *et al.*, 2014; Francisco *et al.*, 2016; Tieman *et al.*, 2017; Xiao *et al.*, 2017). To complement the information on genetic determinants of primary metabolic network obtained from our previous linkage mapping by combining the advantages of association analysis, here we performed gas chromatography–mass spectrometry-based metabolite profiling across four tissues [seedling leaf, mature leaf, young kernel 15 days after pollination (DAP) and mature kernel] of a maize association panel containing 513 inbred lines that were genotyped by 1.25 million genome-wide single nucleotide polymorphisms (SNPs) (Yang *et al.*, 2010; Liu *et al.*, 2017). We conducted GWAS of 61 primary metabolites using this association panel. We also linked the expression level of genes in the kernel at 15 DAP (Fu *et al.*, 2013) with the levels of the identified metabolites in the same tissue using our newly developed linear model, named quantitative genome-wide association study (qGWAS) (Wen *et al.*, 2016b).

To add to the information gathered through GWAS, qGWAS and bioinformatics, we conducted transgenic and candidate gene resequencing approaches to determine the genetic components and variations of the metabolites from the aspartate, aromatic amino acid and trehalose pathways. Nine amino acids (i.e. lysine, methionine, threonine, phenylalanine, tryptophan, valine, isoleucine, leucine and histidine) that are not synthesized by humans and other animals are classified as essential (Galili and Amir, 2013). Fortunately, human and other animals can absorb these

amino acids by eating crops because plants can synthesize them. In the present work, we paid attention to the maize aspartate and shikimate pathways and corroborated the function of the associated genes within maize amino acid metabolism. The aspartate family pathway in plants leads to the biosynthesis of lysine, methionine, threonine and isoleucine (Jander and Joshi, 2009). Aspartate kinase (AK) commits the first catalytic step of this pathway and is adjusted in the negative feedback-loop by several end-products, such as lysine and threonine (Galili, 1995). Lysine is an important essential amino acid nutritionally; its biosynthesis from aspartate-4-semialdehyde is a six-step pathway that is strongly influenced by the rate of its synthesis in plants (Stepansky *et al.*, 2006). Lysine levels in plants are also influenced by its catabolism, whereby lysine is degraded into glutamate and acetyl Co-A. The first two enzymes of lysine catabolism are synthesized from a single *lkr/sdh* gene. Lysine is converted to saccharopine by lysine-ketoglutarate reductase (LKR) through condensation with alpha-ketoglutarate (2OG) and subsequently to glutamate and alpha-amino adipate-delta-semialdehyde by saccharopine dehydrogenase (SDH) (Reyes *et al.*, 2009). The well-tuned lysine metabolism suggests that lysine may serve as a signaling molecule affecting plant growth and interaction with the environment (Galili, 2002; Stepansky *et al.*, 2006). Glutamate, the product of lysine degradation, is decarboxylated by glutamate decarboxylase (GAD) to γ -amino butyric acid (GABA), which is a four-carbon non-protein amino acid present in all organisms (Baum *et al.*, 1996). In humans depression and insomnia are often caused by a large decrease in GABA content as it acts as an inhibitory neurotransmitter (Hall *et al.*, 2011; Bachtiar *et al.*, 2015). In plants, GABA homeostasis is important for plant growth (Takayama and Ezura, 2015). Aromatic amino acids (tryptophan, phenylalanine and tyrosine), which are derived from the shikimate pathway, serve as precursors of a wide range of secondary metabolites (Maeda and Dudareva, 2012). Chorismate, the final product of the shikimate pathway, is converted to prephenate. The conversion of prephenate to phenylalanine and tyrosine may occur via two alternative routes, i.e. the arogenate pathway and the phenylpyruvate or 4-hydroxyphenylpyruvate pathway (Maeda and Dudareva, 2012). In the former pathway, prephenate is first transaminated to arogenate, which is subsequently converted to phenylalanine by arogenate dehydratase (ADT) or to tyrosine by arogenate dehydrogenase (ADH), respectively. In the latter pathway, prephenate is first catalyzed by prephenate dehydratase (PDT) to phenylpyruvate or catalyzed by prephenate dehydrogenase (PDH) to 4-hydroxyphenylpyruvate, followed by transamination of the corresponding products, to phenylalanine or tyrosine, respectively. In plants the arogenate pathway was indicated as the predominant route to phenylalanine biosynthesis (Maeda *et al.*, 2010). A recent study

suggested that alternative tyrosine biosynthesis pathways with distinct routes and localization have evolved in different plants, although their physiological functions remain unknown (Schenck *et al.*, 2017). Trehalose is an alpha, alpha-1,1-linked non-reducing disaccharide (Schluepmann and Paul, 2009). In plants, the trehalose pathway has long been associated with abiotic stress, including drought and excess salt (Henry *et al.*, 2015). Trehalose is derived from UDP-glucose to glucose-6-phosphate and catalyzed by the enzyme trehalose-6-phosphate synthase (TPS). After that, the intermediate trehalose-6-phosphate (T6P) is dephosphorylated into trehalose by the enzyme trehalose-6-phosphate phosphatase (TPP). Subsequently the enzyme trehalase is responsible for the degradation of this metabolite (Zhou *et al.*, 2013). Although some major enzymes in the above-mentioned pathways have been biochemically characterized and the corresponding genes identified in plants, only a handful of genes involved in these pathways were identified by population genetics approaches and limited genetic studies have been performed with plant enzymes. As an increased number of studies have detected considerable loci associated with metabolites in these pathways in plants (Sauvage *et al.*, 2014; Wen *et al.*, 2015; Deng *et al.*, 2017), a large number of responsible genes within these loci and the expression pattern, the subcellular localizations as well as the naturally occurring variation of these genes need to be explored.

A number of significant loci or candidate genes affecting the abundance of primary metabolites in this study have been revealed by GWAS. Moreover, a combination of GWAS, qGWAS, bioinformatics and the analysis of validation transgenic lines enabled us to dissect the key genetic components of maize primary metabolism, thereby facilitating maize improvement.

RESULTS

Metabolic loci identified by GWAS

In total, 61 primary metabolites were identified and quantified across four tissues in a maize association panel (54 in seedling leaf, 52 in leaf at reproductive stage, 47 in young kernel, 55 in mature kernel), 38 of which were measured in all four tissues (Data S1 in the online Supporting Information). All of these metabolites are of known chemical structure and could be assigned to one of the following compound classes: sugars, sugar alcohols, organic acids or amino acids. Detailed information about these metabolites is provided in Data S1. Levels of these 38 metabolites were significantly different across the four types of tissue (Data S1).

A total of 153 significant loci associated with metabolite levels across all four tissues ($P \leq 2.04 \times 10^{-6}$; Figure S1, Data S2) were detected by GWAS. The metabolic variance explained by each locus ranged from 5.0% to 25.2%, with a mean of 7.8%. Only a handful of these 153 loci are known

or well characterized. For instance, opaque-2 (*O2*), which encodes a basic leucine zipper protein transcription factor (Schmidt *et al.*, 1992; Deng *et al.*, 2017), was significantly associated with the level of amino acids leucine, ornithine, phenylalanine, glycine, tyrosine, valine and histidine in the mature kernel in this study (Table 1, Data S2).

Prioritizing candidate genes for primary metabolite variation based on multiple lines of evidence

Prioritizing and refining the signals identified by GWAS and qGWAS is essential for identifying the functional genetic variants and revealing the biological processes underlying the natural variation of maize primary metabolism. Thus, for each of the 153 loci detected here we evaluated the link between functional annotation of all the genes at the locus and the corresponding metabolite to pinpoint the most plausible causal gene. Besides obvious candidates directly involved in the respective pathway, GWAS may identify additional unexpected loci. Therefore, for each of these 153 loci, the candidate genes were identified within a 100-kb window of the lead SNP and are listed in Data S3. If there was no obvious candidate gene, the one nearest to the peak SNP was chosen. Two candidate genes within these 153 significant loci were revealed by qGWAS and 32 exhibited *cis* expression QTLs (Liu *et al.*, 2017) together with significant ($P \leq 0.05$) correlation between gene expression and metabolite level (Data S3).

For qGWAS, we analyzed associations between the expression levels of 28 769 genes (Fu *et al.*, 2013) and the content of 47 metabolites in maize kernels at 15 DAP. Links between 18 metabolites and 760 genes were found by using the REG model (false discovery rate ≤ 0.05). The number of genes whose expression levels significantly relate to the metabolite level ranged from one for glucose to 265 for glycine; and the metabolic variance that the expression of each gene could explain ranged from 3.3% to 12.7%, with a mean of 4.4% (Data S4). There are 15 genes whose expression level was significantly associated with the level of more than two metabolites, including the well-characterized bifunctional enzyme LKR/SDH (Data S4). Significant associations between the level of *lkr/sdh* expression and five metabolites (i.e. glutamine, methionine, isocitric acid, threonine and valine) were revealed (Data S4).

Promising candidate genes are listed in Table 1. All of these genes were supported by multiple lines of evidence, i.e. linkage mapping according to our previous study on primary metabolites (Wen *et al.*, 2015), expression quantitative trait locus (eQTL) analysis, qGWAS or functional annotation that matched with the corresponding metabolite. Although some genes such as *O2* and *lkr/sdh* have been identified previously (Mertz *et al.*, 1964; Kemper *et al.*, 1999), the significant associations between them and multiple metabolites provide us with a more comprehensive picture of their role in metabolic regulation.

Table 1 Summary of selected significant loci

Chr	Candidate gene	Metabolite	Tissue ^a	Annotation	Evidence
1	<i>GRMZM2G124353*</i>	Alanine_beta	l	Alanine:glyoxylate aminotransferase	GWAS, linkage mapping
1	<i>GRMZM2G056469</i>	Dopamine, tyramine	s, l	Tyrosine/DOPA decarboxylase	GWAS, linkage mapping
1	<i>GRMZM2G093125*</i>	Dopamine, tyramine	s, l	Tyrosine/DOPA decarboxylase	GWAS, linkage mapping
1	<i>GRMZM2G017110*</i>	GABA	l	Glutamate decarboxylase	GWAS
1	<i>GRMZM5G826838*</i>	GABA	s	Glutamate decarboxylase	GWAS
1	<i>GRMZM5G825854</i>	Galactinol	k	Sec34-like family protein	GWAS, eQTL
1	<i>GRMZM2G131575</i>	Glutaric_acid_2_oxo	s	H(+)-transporting two-sector ATPase	GWAS
1	<i>GRMZM2G306732*</i>	Inositol_myo	k	Fructose-1,6-bisphosphatase	GWAS
1	<i>GRMZM2G050851*</i>	Leucine	k	Regulation of transcription, DNA-dependent	GWAS, eQTL
1	<i>GRMZM2G046101</i>	Sucrose	l	Glucan endo-1,3-beta-glucosidase	GWAS
1	<i>GRMZM2G077181*</i>	Sucrose	s	Galactinol-sucrose galactosyltransferase	GWAS
1	<i>GRMZM2G162690</i>	Trehalose	k	Trehalase 1	GWAS, eQTL
2	<i>GRMZM2G076204*</i>	Glutamic_acid	k	Gamma-soluble NSF attachment protein	GWAS, eQTL
2	<i>GRMZM2G004590*</i>	Quinic_acid	k, m	Shikimate kinase	GWAS
3	<i>GRMZM2G466833</i>	Caffeic_acid_trans	s	Malate dehydrogenase	GWAS
3	<i>GRMZM2G082780</i>	Galactinol	m	Phosphoenolpyruvate carboxylase	GWAS
3	<i>GRMZM2G054663*</i>	Glyceric_acid	k	D-glycerate 3-kinase	GWAS, eQTL
3	<i>GRMZM2G152127</i>	Glyceric_acid	k	Long-chain-fatty-acid: CoA ligase	GWAS, eQTL
3	<i>GRMZM2G029219*</i>	Pyruvic_acid	k	Carbohydrate transporter	GWAS, eQTL
4	<i>GRMZM2G181362</i>	Gln; Met; isocitric acid; Thr; Val	k	Lysine-ketoglutarate reductase/saccharopine dehydrogenase	qGWAS
5	<i>GRMZM2G121460*</i>	Citric_acid	k	Inosine-5'-monophosphate dehydrogenase	GWAS, eQTL
5	<i>GRMZM2G075265</i>	Glucose	m	ALG6, ALG8 glycosyltransferase family	GWAS
5	<i>GRMZM2G110881</i>	Rhamnose	k, m	UDP-glucose 4-epimerase	GWAS
5	<i>GRMZM2G410865</i>	Rhamnose	k, m	UDP-glucose 4-epimerase	GWAS
5	<i>GRMZM2G076524*</i>	Succinic_acid	m	Succinate dehydrogenase	GWAS
5	<i>AC210013.4_FG017</i>	Sucrose	l	Polygalacturonase	GWAS
6	<i>GRMZM2G124671</i>	Glycerol_3_phosphate	m	Glycosyltransferase family protein	GWAS
6	<i>GRMZM2G027723</i>	Phenylalanine	s	CesA-2	GWAS
6	<i>GRMZM2G442804</i>	Serine	l	S-adenosylmethionine-dependent methyltransferase	GWAS
7	<i>GRMZM2G015534</i>	Leu; Val; Orn; His; Tyr; Phe; Gly	m	Regulatory protein opaque-2	GWAS
7	<i>GRMZM2G121546*</i>	Phenylalanine	k	Arogenate dehydratase	GWAS, qGWAS
7	<i>GRMZM2G342895*</i>	Phenylalanine	k	Arogenate dehydratase	GWAS, qGWAS
7	<i>GRMZM2G113056</i>	Proline	k	Shikimate kinase	GWAS
8	<i>GRMZM2G119345</i>	GABA	k	ABC-2 type transporter family protein	GWAS, eQTL
8	<i>GRMZM2G066024</i>	Sucrose	l	Fructose-bisphosphatealdolase cytoplasmic isozyme	GWAS
10	<i>GRMZM2G400999*</i>	Xylose	l	Xylanase inhibitor protein 1	GWAS

Genes with an asterisk (*) locate closest to the lead SNP.

^aKey: 's' represents seedling leaf, 'l' represents mature leaf, 'k' represents young kernel, 'm' represents mature kernel.

In most cases, the protein or enzyme encoded by the listed candidate gene directly catalyzes the production of the corresponding metabolite; for instance, glutamate decarboxylase catalyzes the production of GABA in mature leaf, succinate dehydrogenase catalyzes the production of succinic acid in the mature kernel, trehalase 1 catalyzes the production of trehalose in the young kernel and finally tyrosine/DOPA decarboxylase catalyzes the formation of tyramine and dopamine in the seedling and mature leaf.

Gene discovery and validation in the amino acid and trehalose metabolic network

The *Ikrt*s (GRMZM5G801369) locus is annotated as lysine ketoglutarate reductase *trans*-splicing related 1. This locus

is significantly associated with the ratio between GABA and 2OG (GABA/2OG, $P = 5.68 \times 10^{-7}$), which are critical components involved in lysine catabolism (Figure 1a, b). Overexpression of *Ikrt*s in maize resulted in a significantly increased level of both lysine ($P = .0031$ in L7342, $P = 0.0128$ in L7298) and the GABA/2OG ratio ($P = .0089$ in L7298) in leaf tissue, which strongly suggests the involvement of *Ikrt*s in the pathway (Figure 1c–e). Moreover, a few other aspartate-derived amino acids are also highly influenced in the over-expression lines; for instance, threonine and GABA are upregulated whereas glutamate is downregulated ($P < .05$; Figure S2 and Figure S3). We also observed a high positive correlation between the expression level of *Zmlkrt*s and the lysine content in over-

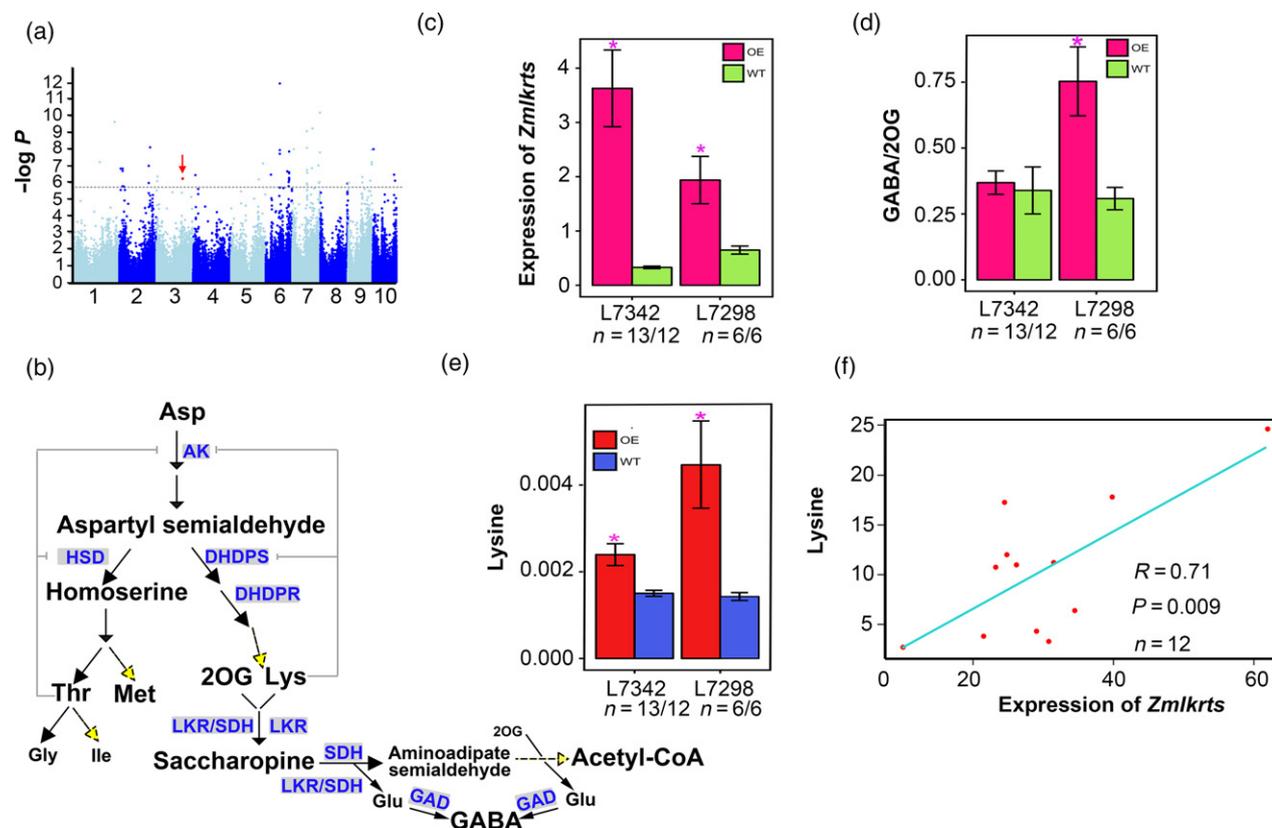


Figure 1. Verification of *Zmlkrts* (lysine ketoglutarate reductase trans-splicing related 1) (*GRMZM5G801369*) as a candidate gene involved in the aspartate-derived amino acid pathway in maize mature leaf.

(a) Manhattan plot displaying the result of the genome-wide association study for the ratio between GABA and alpha-ketoglutarate (2OG) (GABA/2OG) in the mature leaf. The red arrow points to gene *GRMZM2G801369*.

(b) Proposed aspartate metabolic pathway in maize. AK, aspartate kinase; HSD, homoserine dehydrogenase; DHDPS, dihydrodipicolinate synthase; DHDPR, dihydrodipicolinate reductase; LKR, lysine ketoglutarate reductase; SDH, saccharopine dehydrogenase; GAD, glutamate decarboxylase. The solid arrows indicate exact enzymatic steps and the yellow dotted arrows indicate reactions with more than one step.

(c) Relative expression level of *GRMZM2G801369* in the leaves of over-expression maize lines (T_1). Values represent mean \pm SEM ($n \geq 6$ plants). * $P < .05$.

(d) The level of GABA/2OG between the wild-type individuals (WT) and the over-expression individuals (OE) in two transgenic lines (L7342 and L7298). Values represent mean \pm SEM ($n \geq 6$ plants). * $P < .05$.

(e) Relative intensity of lysine between WT and OE individuals in two transgenic lines (L7342 and L7298). Values represent mean \pm SEM ($n \geq 6$ plants). * $P < .05$.

(f) Correlation between the expression level of *GRMZM5G801369* and lysine content in leaves of over-expression rice lines (T_0).

expression lines of rice ($r = .71$, $P = .0097$; Figure 1f). Taken together, we thus propose that *Zmlkrts* plays a key role in the regulation of aspartate levels via its role in the lysine catabolic pathway.

The level of GABA in mature maize leaves was mapped to a putative glutamate decarboxylase (*ZmGAD1*; GRMZM2G017110) by GWAS (Figure 2a, b). Overexpression of *ZmGAD1* in rice led to the accumulation of GABA and reduction of glutamate in the leaf tissue (Figure 2c–e). The regulatory influence of *ZmGAD1* on GABA and glutamate level in rice leaf strongly supports its functional annotation as ‘glutamate decarboxylase’. In addition, a negative correlation between glutamate content and the expression level of *ZmGAD1* ($r = -.19$, $n = 213$, $P = .0066$) in the mature leaves was found in our maize association panel.

A locus (*ZmADT*) was identified on chromosome 7 that is strongly associated with the level of phenylalanine in the young kernel (kernel at 15 DAP) (Figure 3a). Arogenate dehydratase catalyzes the last step of this pathway linking arogenate to phenylalanine (Figure 3c) (Holding *et al.*, 2010). Two tandem genes (*GRMZM2G342895* and *GRMZM2G121546*) both annotated as arogenate dehydratase (*ADT*; B73 RefGen v2) co-locate with this locus (Figure 3b). In addition, significant association between the level of phenylalanine in the young kernel and both *GRMZM2G342895* and *GRMZM2G121546* was identified by qGWAS and Pearson correlation analysis (Figure 3f). By resequencing this locus we found a polymorphism (811-bp insertion–deletion, InDel 811) at the 5' untranslated region (UTR) of *GRMZM2G342895*, and the allele frequency of the 811-bp insertion is 46% (Figure 3d). The 811-bp insertion

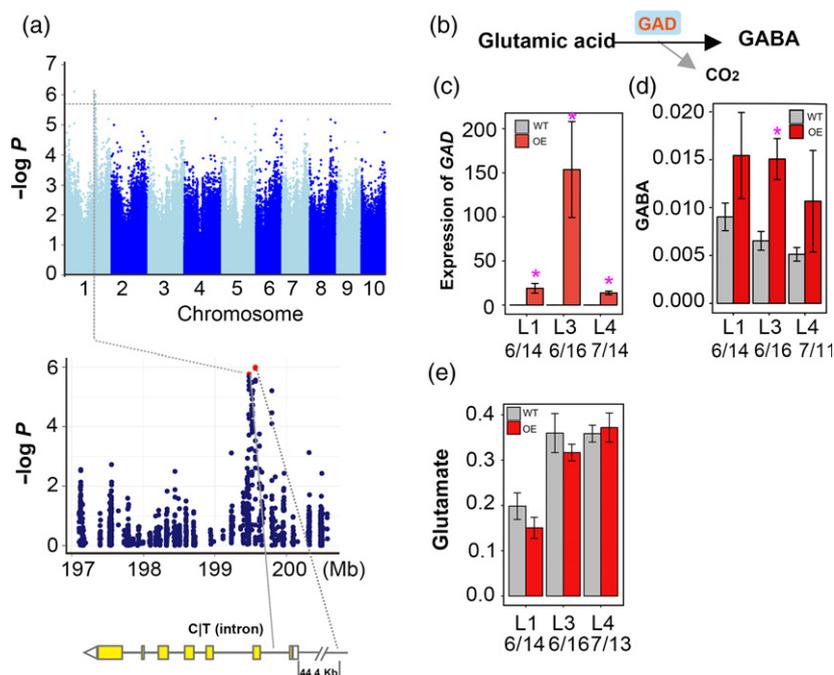


Figure 2. Verification of *ZmGAD1* (*GRMZM2G017110*) as a candidate gene responsible for the level of GABA in mature maize leaf.

(a) Manhattan plot displaying the result of the genome-wide association study for GABA content in the maize mature leaf. The dashed line indicates the threshold of $P \leq 2.04 \times 10^{-6}$. The lower panel shows the regional association between single nucleotide polymorphisms (SNPs) and GABA level (100 kb upstream and downstream of the most significant SNP) and the relative position of the peak SNP and the corresponding gene.

(b) The reaction from glutamate to GABA catalyzed by glutamate decarboxylase (GAD).

(c) Expression profiling of *ZmGAD1* in the leaves of rice transgenic lines L1, L3 and L4 (T_1). Values represent mean \pm SEM ($n \geq 6$ plants). * $P < .05$.

(d) Relative intensity of GABA in wild-type individuals (WT) and over-expression individuals (OE) of transgenic lines L1, L3 and L4. Values represent mean \pm SEM ($n \geq 6$ plants). * $P < .05$.

(e) Relative intensity of glutamate in WT and OE individuals of transgenic lines L1, L3 and L4. Values represent mean \pm SEM ($n \geq 6$ plants).

strongly decreases the expression of *GRMZM2G342895* ($P = 1.08 \times 10^{-7}$; Figure 3e) and the phenylalanine content in the young kernel ($P = 5.1 \times 10^{-4}$; Figure 3e), which may represent one of the key functional genetic variants for the *ZmADT* locus.

In this study, a locus (*Tre1*) on chromosome 1 was significantly associated with the level of trehalose in the maize young kernels (Figure 4a, c; $P = 1.95 \times 10^{-7}$, mix linear model (MLM)). Three SNPs which are in a linkage disequilibrium (LD) block were identified above the threshold ($P \leq 2.04 \times 10^{-6}$, MLM). One of these three SNPs was located in the third exon of *Tre1*, resulting in an amino acid replacement (from serine to glycine; Figure 4b). A *cis*-eQTL was identified for the expression level of *ZmTre1* (Figure 4d) and negative correlation ($r = -.25$, $P = 8.07 \times 10^{-4}$ by Pearson correlation analysis) was observed between the *Tre1* expression level and the level of trehalose in the young kernel (Figure 4e). We therefore resequenced the promoter region of *Tre1* and found three insertion–deletion (InDel) polymorphisms, which are located within 500 bp (Figure 4b). The trehalose levels in the inbred lines with insertion (with any one of the three insertions or any two or all three insertions) in the upstream of *ZmTre1* are significantly ($P = 1.4 \times 10^{-5}$) higher than those with no

insertions (Figure 4f). Also, *ZmTre1* expression is higher in lines with insertion than lines without (*t*-test, $P = 6.9 \times 10^{-13}$) (Figure 4f). In addition, we conducted haplotype analysis by combining InDel 1_1 with the three significant SNPs, respectively. Significant differences between distinct haplotypes were found when InDel 1_1 was combined with SNP chr1.S_79999858 (Figure 4g). We thus speculate that the three InDels, SNP chr1.S_79999858 and its combinations, are all responsible for the trehalose level in the young kernel.

DISCUSSION

Understanding the genetic control of primary metabolism in maize, the crop with the largest production worldwide, is essential for worldwide food security and sustainable agriculture. Comprehensive studies on the natural genetic variation of maize primary metabolism are scarce despite exhaustive studies that focus on only one or a few metabolic traits in maize (Zhang *et al.*, 2015; Deng *et al.*, 2017). This study measured primary metabolites in multiple tissues of maize and demonstrated that there were few loci with a large effect (with $R^2 > 15\%$); however, for the variation of primary metabolites loci with a minor to modest genetic contribution are common. Each locus contributes a

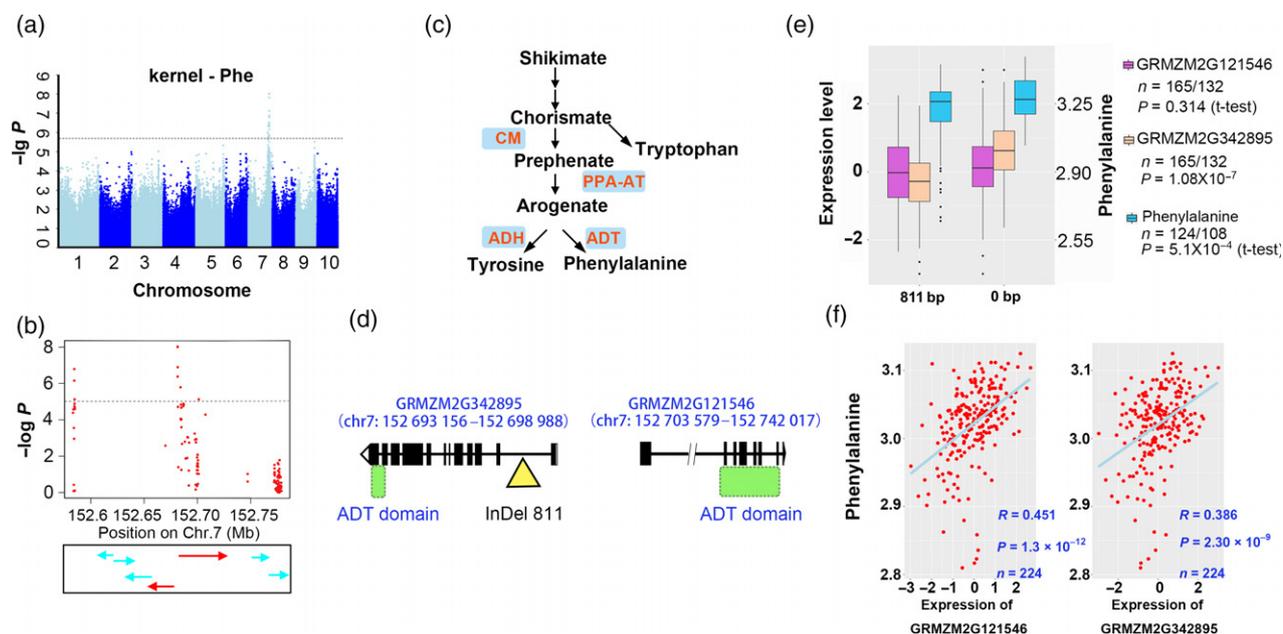


Figure 3. Resequencing and candidate gene association analysis of *ZmADT* as a candidate gene affecting phenylalanine content in young maize kernels. (a) Manhattan plot displaying the result of the genome-wide association study for phenylalanine level in the young maize kernel. (b) Regional association between single nucleotide polymorphisms (SNPs) and phenylalanine level (100 kb upstream and downstream of the most significant SNP and genes annotated in this region). Genes are represented by arrows and the red arrows indicate two genes at the locus *ZmADT*. (c) Proposed shikimate metabolic pathway in maize. CM, chorismate mutase; PPA-AT, prephenate amino transferase; ADH, arogenate dehydrogenase; ADT, arogenate dehydratase. (d) Structure of the two *ZmADT* genes. The green boxes indicate the ADT domain. The position of the 811 insertion is indicated by the yellow triangle. (e) Box plot showing differences in phenylalanine content and expression level of two *ZmADT* genes between the two genotypes at the site InDel 811. (f) Plot of Pearson correlation between the content of phenylalanine and the normalized expression level of the two *ZmADT* genes.

modest effect on average ($R^2 = 7.8\%$). Genome-wide association of carbon and nitrogen metabolism in the maize nested association mapping population was revealed by Zhang *et al.* (2015), who identified the genetic basis of natural variation controlling the levels of 12 key C and N metabolites. It was also indicated that most QTLs of the variation of C and N metabolites have modest effects. Although the contribution of individual primary metabolic QTLs to the total phenotypic variance (PVE) usually has a wide range, for example the PVE of each QTL identified in an Arabidopsis RIL population ranged from 1.7% to more than 52.1% (Liseic *et al.*, 2008) and the PVE of each QTL identified in a maize RIL population ranged from 2.4% to 49.0% (Wen *et al.*, 2015), the average contribution of individual primary metabolic QTLs is generally modest (Liseic *et al.*, 2008; Sauvage *et al.*, 2014; Wen *et al.*, 2015). The relative low level of primary metabolic variation explained by each locus can be due to genes and metabolites being embedded in a complex regulatory network of primary metabolism. Although a complex genetic basis underlying the variation of primary metabolite levels has been revealed by previous studies and in this study (Sauvage *et al.*, 2014; Alseikh *et al.*, 2015; Wen *et al.*, 2015; Zhang *et al.*, 2015), GWAS and qGWAS could efficiently fine-map the potential causative loci, and in some cases help

identify the casual genes. Consistent with previous studies the majority of loci identified by GWAS here were tissue-specific, which implies the presence of distinct genetic and biochemical regulation of the pathways concerned (Wen *et al.*, 2015). Since different tissues use only a subset of the capabilities encoded by the maize genome and spatio-temporal differences in the expression patterns of genes also lead to metabolic diversity among different tissues, dissecting metabolites in multiple tissues provided more comprehensive information than that found in a single tissue. The identification of hundreds of loci in the present multiple-tissue primary metabolite GWAS enables candidate gene identification, will aid further analysis of the molecular basis of variation of maize primary metabolites and provides a foundation on which to design breeding strategies for metabolic engineering. The generally low level of explained phenotypic variation estimated here implies that significant improvement may require the pyramiding of several loci.

By integrating gene expression information we can make the connection between the metabolite concentration and gene expression within the same tissue and predict the most likely genes and genetic variants responsible for primary metabolic pathways, as well as indicate potential targets for transgenic manipulation. The observed genetic

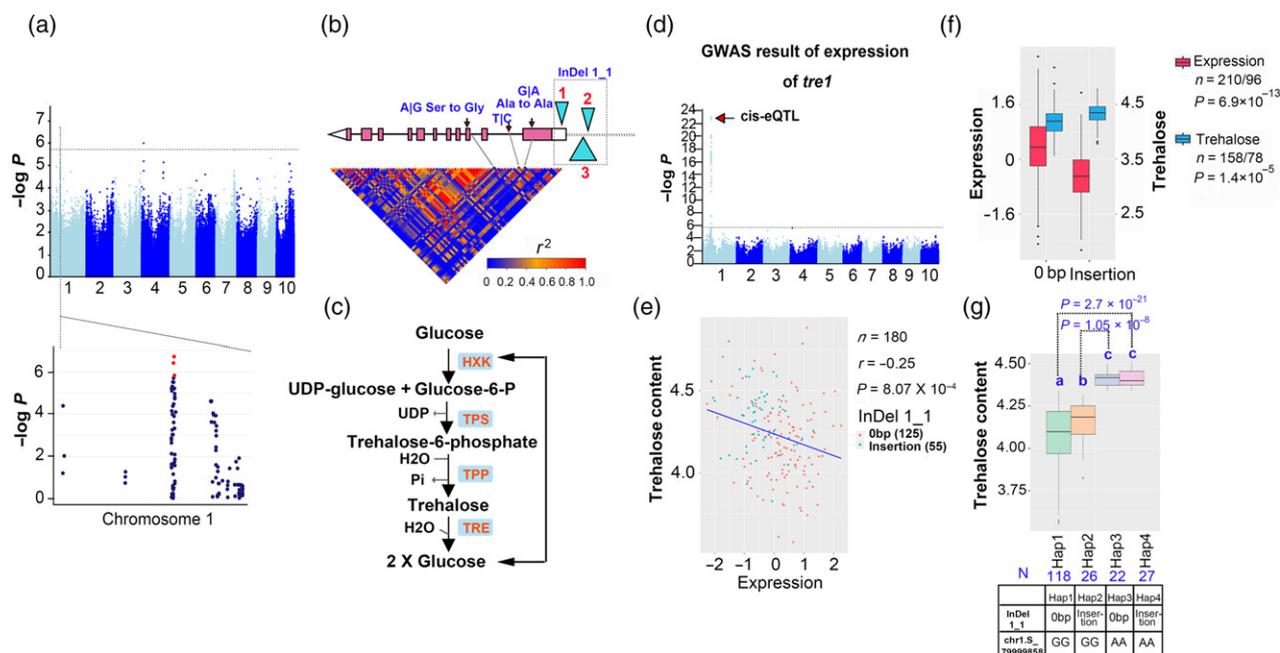


Figure 4. Resequencing and candidate gene association analysis of *ZmTre1* as a candidate gene affecting trehalose content in maize young kernel.

(a) Manhattan plot displaying the result of the genome-wide association study for trehalose in maize young kernel (upper), and the regional association plot for locus *ZmTre1* (lower). Red dots indicate the single nucleotide polymorphisms (SNPs) above the threshold. (b) Structure and genetic variations of the candidate gene, *ZmTre1*. The black arrows point to the relative positions of the three significant SNPs in (a), with blue words denoting the alleles and amino acid replacements. The B73 allele on reverse strand is before the vertical line. The aquamarine triangles indicate InDel 1_1 with three different fragments inserted into the promoter region marked by the dashed black box. Below the gene structure is the linkage disequilibrium plot showing the pair-wise r^2 value among all polymorphic sites in *ZmTre1* and the asterisks on it indicate the three significant SNPs which are connected to the three black arrows in the gene. (c) Proposed trehalose metabolic pathway in maize. HXK, hexokinase; TPS, trehalose-6-phosphate synthase; TPP, trehalose-6-phosphate phosphatase; TRE, trehalase; Glucose-6-P, glucose-6-phosphate; UDP, uridine-5-diphosphate; Pi, phosphate. (d) Manhattan plot of eQTL analysis of *ZmTre1* expression. The red arrow indicates the *cis*-expression quantitative trait locus (*cis*-eQTL) signal of *ZmTre1*. (e) Plot of Pearson correlation between the content of trehalose and the normalized expression level of the *ZmTre1* gene. Maize inbred lines with different genotypes at the InDel 1_1 site are shown in orange and midnight blue, respectively. (f) Box plot for trehalose content and *ZmTre1* expression level between different genotypes of InDel 1_1. (g) Box plot showing the difference in trehalose content between different haplotypes.

variation at a locus is usually supported by *cis*-eQTL of the target gene and significant correlation between gene expression and metabolite levels (Wen *et al.*, 2014). The causal variants can have regulatory consequences which influence the metabolite level through transcriptional regulation. *A priori* knowledge of the structural and regulatory genes participating in plant primary metabolism helped us select potential candidates in this study. For instance, *ZmGAD1* and *ZmADT*, which we validated here, putatively encode enzymes that catalyze the production of the target metabolites used in GWAS. However, genes of unknown function that were found by using expression data are also candidates for further investigation. Although qGWAS did find well-characterized functional genes and there is overlap between the GWAS and qGWAS result in this study, not many genes were cross-validated using these two approaches. This may be due to following reasons. First, the influences on the metabolite level from the genetic variants and from the transcriptional variants can be independent. Second, environmental factors may result in

disagreement since the expression and the metabolite data were collected in samples grown in different environments.

Free amino acids serve as essential building blocks for proteins and are important for the biosynthesis of numerous essential compounds and responses to environmental changes during normal plant growth and development (Tzin and Galili, 2010; Angelovici *et al.*, 2011). Although maize is the highest-yielding crop worldwide, deficiency in the essential amino acids (i.e. lysine and tryptophan) in its main storage proteins, the zeins, prevents it from being the sole protein source for humans and livestock. The discovery of the *O2* mutation was of considerable importance for improvement of maize protein quality by enhancing the lysine and tryptophan levels while decreasing the synthesis of zeins, and in compensation increasing other (non-zein) seed proteins (Liu *et al.*, 2017). The significant association between the *O2* locus and multiple amino acid levels identified by GWAS here implies an opportunity for mining naturally occurring genetic variants for maize quality improvement or manipulation of primary metabolism. In

addition, the importance of a known locus in the lysine catabolic pathway (*lkr/sdh*) was indicated in this study, providing a proof-of-concept for our approach. However, the identification of *lkr/sdh* is of greater significance than this alone. While transgenics in which *lkr/sdh* was repressed in combination with upregulated *AK* (aspartate kinase) and *dhdps* (dihydrodipicolinate synthase) expression were documented to have considerably higher lysine levels; they also exhibited aberrant germination in Arabidopsis and soybean and have a large impact on the physicochemical characteristics of grain flours in rice (Galili and Amir, 2013; Yang *et al.*, 2016). Our results suggest that the use of absolute metabolite levels as target traits that can reveal gene and metabolite association is an efficient method with considerable power. Moreover, using a population-based approach allows immediate screening for yield penalties or other physiological disadvantages. However, when dynamic and unknown connections between the components exist in a metabolic network, the sum or the ratio of the related traits can also serve as a derived trait for genetic mapping (Angelovici *et al.*, 2013, 2017).

Herein we used the ratio between two components of lysine catabolism as a trait (GABA/2OG) and found locus *Zmlkrt1*. Three genes (*AC234165.1_FG002*, *GRMZM5G801369* and *AC234165.1_FG003*) are located at this locus, and the former two are annotated as lipoprotein and lysine ketoglutarate reductase *trans*-splicing related, respectively. Functional annotation of *AC234165.1_FG003* is unknown. We thus consider the putative function of *Zmlkrt1* as related to lysine ketoglutarate reductase *trans*-splicing. Our current validation through overexpression of *Zmlkrt1* in maize strongly supports its regulatory influence within the pathway. A recent study integrating omics networks in a developmental atlas of maize revealed that *Zmlkrt1* is related to RNA metabolism (Walley *et al.*, 2016). However, more work is required to resolve the molecular and biochemical influence of this locus on the lysine catabolic pathway and to understand the exact molecular mechanism. The proteogenomics approach may represent a route to achieve this, given that it readily provides information on the frequency of translation of the different splice variants (Zhu *et al.*, 2017). Whilst particularly pertinent in this instance it may also uncover alternative splicing at other QTLs.

At the locus *ZmGAD1*, associated here with the level of GABA in mature maize leaves, a candidate gene putatively encoding a glutamate decarboxylase is of higher priority for selection. In plants GABA is not only a metabolite but also a signaling molecule, and is involved in a variety of physiological processes such as pollen tube growth (Yu *et al.*, 2014), regulation of intracellular Ca²⁺ levels (Fait *et al.*, 2008) and ethylene production (Shi *et al.*, 2010; Takayama *et al.*, 2017). The maize genome encodes five putative glutamate decarboxylase (*GAD*) genes and their

expression pattern greatly varies (Sekhon *et al.*, 2011). One of them is specifically expressed in the anther, and two others, whilst being expressed in multiple tissues, have a dramatically higher expression level in the primary root and silk, respectively. *ZmGAD1* identified here has a higher expression level in mature leaves and internodes than that in other tissues (Sekhon *et al.*, 2011).

Potential genetic changes within *Zmlkrt1* and *ZmGAD1* were detected by resequencing their 5' and 3' UTRs and part of the coding regions (Figures S4 and Figure S5). For *lkrts*, a 165-bp insertion in the 5' UTR, which is a DNA transposon named hobo-activator based on database via RepeatMasker, was found in only six lines and might be functional (Figure S4b, c). However, it is not reasonable to judge the significance of this insertion and phenotypic variance due to the low allelic frequency (Figure S4d). For *GAD1*, the 3' UTRs of 443 varieties were sequenced and a 71-bp insertion in the last intron of *ZmGAD* (*ZmGAD_T03*) was detected in eight lines (Figure S5a–c). It may affect the abundance of different transcripts of *ZmGAD* based on our RNA sequencing (RNA-seq) data; however, it is not reasonable to judge the significance of this insertion and phenotypic variance due to the low allelic frequency (Figure S5d).

One of the promising genes identified, *ZmADT* (*GRMZM2G342985*; B73 RefGen v2) annotated to encode ADT, co-locates with a locus significantly associated with the level of phenylalanine in the young kernel (Table 1). The expression level of *ZmADT* is also highly linked to the level of phenylalanine, as indicated by qGWAS. The enzyme ADT participates in phenylalanine biosynthesis by catalyzing the conversion of arogonate into phenylalanine, H₂O and CO₂. Aromatic amino acid biosynthesis provides basic building modules for proteins and diverse metabolites in plants (Maeda and Dudareva, 2012), and hence this information will be important for the breeding of nutritionally improved crops.

Trehalose and its precursor T6P have been found to possess a variety of different functions, including storage of chemical energy and biotic and abiotic stress tolerance (Henry *et al.*, 2014; Bledsoe *et al.*, 2017). Recently, Nuccio *et al.* (2015) pointed out that engineering of trehalose metabolism may provide a good target for future crop improvement. The significant role of the trehalose pathway can be revealed by dramatic phenotypes of plants with altered expression of trehalose pathway genes, which include effects on diverse agronomically important traits (Lunn *et al.*, 2014). However, the detailed picture of the function of the trehalose pathway and genes involved in this pathway in a C₄ cereal crop such as maize remain unclear. In total, 14 trehalose-6-phosphate synthase (TPS) genes, 11 trehalose-6-phosphate phosphatase (TPP) genes and one trehalase (TRE) gene have been identified in the maize genome (Henry *et al.*, 2014). Herein, through GWAS, candidate gene resequencing and association analysis, we

have identified the TRE gene and speculated on the genetic and regulatory underpinnings of the naturally occurring variation of the trehalose level in the maize kernel at 15 DAP. Although a more detailed characterization of *ZmTre1* is necessary for further understanding whether or how it plays a role in sugar and starch metabolism, plant growth and/or stress tolerance, the association panel used here offers us a fascinating system for investigating the naturally occurring variation of the trehalose pathway. In addition, it is worth dissecting genetic variants of the maize family of TPS and TPP genes using this available association panel and assessing contributions of these genes to the maize metabolism and plant performance in the context of natural variation.

In summary, genetic variation within our association panel represents a rich resource for maize improvement. The combination of a GWAS with metabolomic, transcriptomic and transgenic approaches enables dissection of complex quantitative traits into their structural and regulatory genetic components, which provides an effective approach for the metabolic engineering of nutritional traits and facilitates the use of existing genetic variations by marker-assisted breeding.

EXPERIMENTAL PROCEDURES

Plant materials

In this study samples for metabolite profiling were taken in an association panel containing 513 diverse inbred lines (Yang *et al.*, 2010; Wen *et al.*, 2015). This panel was planted in Wuhan (114°21' E, 30°28' N) in spring 2013 and Sanya (109°51' E, 18°25' N) in winter 2013. For each line planted at the field station in Wuhan, we took samples from three different stages, that is, leaf at seedling stage, leaf at reproductive stage and kernels at 15 DAP. The two leaf stages belong to stage code 17 and code 61 ([http://research.omicsgroup.org/index.php/BBCB-scale_\(maize\)](http://research.omicsgroup.org/index.php/BBCB-scale_(maize))), respectively. At the field station in Sanya we took samples of mature kernels of each line.

For leaf samples, leaves of three plants from the same line were collected and bulked as one. For kernel samples, kernels from three ears from the same line were collected and bulked as one. Normally we expected to have around 510 samples; however, for some genotypes we failed to harvest three cobs/ears (e.g. some ears contain fewer than 15 well-pollinated kernels). To reduce bias we did not use the data from these genotypes. In total, 1256 samples were collected in Wuhan (468 samples from leaf at seedling stage, 493 samples from leaf at reproductive stage, and 295 samples from 15-DAP kernel) and 490 samples were collected in Sanya (mature kernel). All samples were collected and stored at -80°C before metabolite extraction.

Genotypic and transcriptomic data

Previously, the panel was genotyped using multiple platforms (i.e. RNA-seq, Genotyping-by-sequencing, SNP array) and genotypes at 2.65 million loci were obtained (Liu *et al.*, 2017). In this study we used 1.25 million out of the 2.65 million loci which had a Minor Allele Frequency (MAF) $\geq 5\%$. The RNA-seq data covering 28 679 genes from 15-DAP seeds of 368 lines of the panel are publicly available (Fu *et al.*, 2013).

Metabolite profiling based on gas chromatography time-of-flight mass spectrometry (GC-TOF-MS)

A protocol adapted from previous studies (Roessner *et al.*, 2001; Liseč *et al.*, 2006) was adopted for metabolite extraction. Briefly, 50 mg of fresh powder of each sample was used for follow-up extraction, as described in detail in our previous publication (Wen *et al.*, 2015). The dried extracted samples were shipped to the Max-Planck Institute of Molecular Plant Physiology (Potsdam, Germany) for metabolite profiling. After derivatization, 1 μ l of each sample was injected into a GC-TOF-MS system [Pegasus III (Leco, <https://www.leco.com/>) for samples collected in Wuhan and Pegasus IV (Leco) for mature kernel samples collected in Sanya]. Gas chromatography was performed and mass spectra were evaluated (Kopka *et al.*, 2005; Wen *et al.*, 2015). The details of each identified metabolite as well as the raw phenotypic data for all lines are provided in Data S5.

Genome-wide association analysis for metabolites detected across four types of tissue

The association between the genome-wide SNPs (about 1.25 million SNPs with MAF $\geq 5\%$) and each metabolic trait detected from different tissue types (including metabolite concentration and selected metabolite/metabolite ratios) was tested. We used a mixed linear model accounting for the population structure (*Q*) and familial relationship (*K*) implemented in TASSEL3.0 software (Zhang *et al.*, 2010). The effective number of independent markers (*N*) was calculated using the GEC software tool (Li *et al.*, 2012), and a suggestive *P*-value threshold ($1/N$) was set to control the genome-wide Type I error rate. The *P*-value threshold was 2.04×10^{-6} for the entire population. The following steps were used to identify significant loci. First, all significantly associated SNPs were grouped into clusters where the distance between two consecutive SNPs was < 20 kb and the clusters with at least two significant SNPs were regarded as candidate loci. Next, those candidate loci in LD ($r^2 \geq .1$) with other more significant candidates for the same trait were considered as false-positive associations introduced by intrinsic LD structure and were thus removed. The SNP with the lowest *P*-value was selected as the lead SNP of a locus that was significantly associated with any metabolic trait. The most probable candidate gene in each locus was selected within 100 kb of the lead SNP. To identify genomic variants responsible for the expression level of the candidate genes associated with the metabolic traits, and to determine whether these significant loci might regulate the transcription of other genes, we performed association analysis as described above using the expression level of each gene as the phenotype. Results of eQTL analysis of candidate genes were obtained from our previous study, where genomic variants responsible for the expression level of each gene detected by RNA-seq on the 15-DAP kernel of 368 maize inbred lines was identified (Fu *et al.*, 2013; Liu *et al.*, 2017). The expression level of 28 769 genes was quantified by our previous RNA-seq study on the 15-DAP kernel of 368 maize inbred lines.

Association of transcriptional and metabolic levels

A statistical method was adopted to link the expression and metabolite levels detected in the young and mature kernels. A naïve linear regression (i.e. the REG model) was fitted that treated each metabolite as a response variable and the gene expression levels as explanatory variables, hereafter referred as 'qGWAS' (Wen *et al.*, 2016b). The naïve REG model was conducted using the R/lm package.

Haplotype analysis

We conducted haplotype analysis by combining InDel 1_1 (0 bp/insertion) with the SNP (chr1.S_79999858). An ANOVA analysis was conducted to test the difference between the four groups of haplotype. The significance threshold is $P < .05$.

Plasmid construction and maize transformation

The coding sequences of candidate genes were amplified from B73. After Sanger sequencing to ensure the correct sequence, the fragment was inserted into the pZZ-EGFP vector by recombination using a ClonExpress II One Step Cloning Kit (Vazyme, <http://en.vazyme.com/>). The pZZ-EGFP vector contains the maize Ubi promoter with an EGFP sequence fused before the Tnos sequence, which was derived from the maize transformation vector pZZ01523 provided by China National Seed Group Co., Ltd (<http://www.chinaseeds.com.cn/>). Confirmed clones were transferred to *Agrobacterium* EHA105 and transformed into maize inbred line C01. To determine whether the fragment was transferred into the genome, the marker gene *Bar* and the expression of the target gene were examined by PCR and qPCR, respectively. Primers are listed in Table S1. The T₀ transgene-positive plants were backcrossed to C01 to obtain T₁ generation. By planting the T₁ individuals from one ear (a segregating line), we identified transgene-positive individuals (over-expression individuals; OE) and transgene-negative individuals (wild type, WT), which were distinguished by Bar-test paper and qPCR analysis on each plant. Metabolite profiling was done to compare between OE and WT offspring of each segregating line.

Plasmid construction and rice transformation

The full length-cDNA-fragments of candidate genes were amplified from B73 and ligated into the vector pCAMBIA1300s. Confirmed clones were introduced into *Agrobacterium* EHA105 by electroporation and calli induced from mature seeds of an elite japonica rice cultivar Zhonghua11. The sequences of marker gene *hpt* and the target gene were detected by PCR, and expression of the target gene was detected by qPCR. Primers are listed in Table S1.

Expression quantification

Total RNA was extracted by TRIzol (Invitrogen, <http://www.invitrogen.com/>) from at least five seedlings of each genotype, according to the manufacturer's instructions. First-strand cDNA was synthesized using a TransScript One-Step gDNA Removal and cDNA Synthesis SuperMix (TransGene Biotek, <http://www.transgenebiotek.com/>) according to the manufacturer's protocol. Quantitative PCR was performed on an optical 96-well plate in a Bio-Rad PCR system (CFX96; <http://www.bio-rad.com/>) with SYBR Mix (Vazyme). The relative expression level was calculated using *ZmAct1* (*GRMZM2G126010*) as an internal control in maize or rice actin (*LOC_Os03.g50885*) as a control in rice. The relative quantification method was used to calculate the expression level (Livak and Schmittgen, 2001). The PCR conditions consisted of an initial denaturation step at 95°C for 3 min, followed by 40 cycles at 95°C for 10 sec, then 58°C for 30 sec and 72°C for 15 sec.

AUTHOR CONTRIBUTIONS

WW, ARF and JY designed and supervised this study. WW, MJ, KL, SA, WL, MZ and FL performed the experiments. WW, MJ, HL and YX performed the data analysis. YB and

LW provided technical support and advice for this study. WW, MJ, ARF and JY prepared the manuscript with inputs from other authors.

ACKNOWLEDGEMENTS

We would like to thank Dr Yongjun Lin for help with rice transformation. This research was funded by the National Key Research and Development Program of China (2016YFD0101003), the National Natural Science Foundation of China (31525017) and the Max Planck Society and Hubei provincial foreign science and technology cooperation project (2017AHB038).

CONFLICTS OF INTEREST

The authors declare no conflicts of interest.

SUPPORTING INFORMATION

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

Figure S1. The distribution of significant primary metabolic loci in the maize genome.

Figure S2. Metabolite profile of *Zmlkrts* over-expression maize lines (T₀).

Figure S3. Fold change (log₂-transformed) of metabolites in over-expression individuals to wild type in each line (T₁).

Figure S4. A 165-bp indel exists in candidate gene *Zmlkrts*.

Figure S5. A 71-bp indel exists in candidate gene *ZmGAD1*.

Table S1. Information on primers used in this study.

Data S1. List of metabolites identified in different tissues of the association panel.

Data S2. List of significant loci for metabolite concentrations identified in different types of tissues of the association panel.

Data S3. List of candidate genes at the significant loci.

Data S4. Summary of significant association between gene expression level and metabolic traits in maize kernel.

Data S5. Data on metabolite level in four tissues of each line.

REFERENCES

- Alonso-Blanco, C., Aarts, M.G., Bentsink, L., Keurentjes, J.J., Reymond, M., Vreugdenhil, D. and Koornneef, M. (2009) What has natural variation taught us about plant development, physiology, and adaptation? *Plant Cell*, **21**, 1877–1896. <https://doi.org/10.1105/tpc.109.068114>
- Alseekh, S., Tohge, T., Wendenberg, R. et al. (2015) Identification and mode of inheritance of quantitative trait loci for secondary metabolite abundance in tomato. *Plant Cell*, **27**, 485–512. <https://doi.org/10.1105/tpc.114.132266>
- Angelovici, R., Fait, A., Fernie, A.R. and Galili, G. (2011) A seed high-lysine trait is negatively associated with the TCA cycle and slows down Arabidopsis seed germination. *New Phytol.*, **189**, 148–159. <https://doi.org/10.1111/j.1469-8137.2010.03478.x>
- Angelovici, R., Lipka, A.E., Deason, N., Gonzalez-Jorge, S., Lin, H., Cepela, J., Buell, R., Gore, M.A. and Dellapenna, D. (2013) Genome-wide analysis of branched-chain amino acid levels in Arabidopsis seeds. *Plant Cell*, **25**, 4827–4843. <https://doi.org/10.1105/tpc.113.119370>
- Angelovici, R., Batushansky, A., Deason, N., Gonzalez-Jorge, S., Gore, M.A., Fait, A. and DellaPenna, D. (2017) Network-Guided GWAS Improves Identification of Genes Affecting Free Amino Acids. *Plant Physiol.*, **173**, 872–886. <https://doi.org/10.1104/pp.16.01287>
- Bachtiar, V., Near, J., Johansen-Berg, H. and Stagg, C.J. (2015) Modulation of GABA and resting state functional connectivity by transcranial direct current stimulation. *eLife*, **4**, e08789.
- Baum, G., Lev-Yadun, S., Fridmann, Y., Arazi, T., Katsnelson, H., Zik, M. and Fromm, H. (1996) Calmodulin binding to glutamate decarboxylase is required for regulation of glutamate and GABA metabolism and normal development in plants. *EMBO J.*, **15**, 2988–2996.

- Bledsoe, S.W., Henry, C., Griffiths, C.A., Paul, M.J., Feil, R., Lunn, J.E., Stitt, M. and Lagrimini, L.M. (2017) The role of Tre6P and SnRK1 in maize early kernel development and events leading to stress-induced kernel abortion. *BMC Plant Biol.*, **14**, 74. <https://doi.org/10.1186/s12870-017-1018-2>
- 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 Modelling Approach to Link Leaf Physiology to Kernel Yield. *Plant Cell*, **29**, 919–943.
- Chan, E.K., Rowe, H.C., Corwin, J.A., Joseph, B. and Kliebenstein, D.J. (2011) Combining genome-wide association mapping and transcriptional networks to identify novel genes controlling glucosinolates in Arabidopsis thaliana. *PLoS Biol.*, **9**, e1001125. <https://doi.org/10.1371/journal.pbio.1001125>
- Deng, M., Li, D., Luo, J., Xiao, Y., Liu, H., Pan, Q., Zhang, X., Jin, M., Zhao, M. and Yan, J. (2017) The genetic architecture of amino acids dissection by association and linkage analysis in maize. *Plant Biotechnol. J.*, **15**, 1250–1263. <https://doi.org/10.1111/pbi.12712>
- Fait, A., Fromm, H., Walter, D., Galili, G. and Fernie, A.R. (2008) Highway or byway: The metabolic role of the GABA shunt in plants. *Trends Plant Sci.*, **13**, 14–19. <https://doi.org/10.1016/j.tplants.2007.10.005>
- Fernie, A.R. and Schauer, N. (2009) Metabolomics-assisted breeding: A viable option for crop improvement? *Trends Genet.*, **25**, 39–48. <https://doi.org/10.1016/j.tig.2008.10.010>
- Francisco, M., Joseph, B., Caligagan, H., Li, B., Corwin, J.A., Lin, C., Kerwin, R.E., Burow, M. and Kliebenstein, D.J. (2016) Genome Wide Association Mapping in Arabidopsis thaliana Identifies Novel Genes Involved in Linking Allyl Glucosinolate to Altered Biomass and Defense. *Front Plant Sci.*, **7**, 1010.
- Fu, J., Cheng, Y., Linghu, J. *et al.* (2013) RNA sequencing reveals the complex regulatory network in the maize kernel. *Nat. Commun.*, **4**, 2832.
- Fukushima, A. and Kusano, M. (2013) Recent progress in the development of metabolome databases for plant systems biology. *Front Plant Sci.*, **4**, 73.
- Galili, G. (1995) Regulation of Lysine and Threonine Synthesis. *Plant Cell*, **7**, 899–906. <https://doi.org/10.1105/tpc.7.7.899>
- Galili, G. (2002) New insights into the regulation and functional significance of lysine metabolism in plants. *Annu. Rev. Plant Biol.*, **53**, 27–43.
- Galili, G. and Amir, R. (2013) Fortifying plants with the essential amino acids lysine and methionine to improve nutritional quality. *Plant Biotechnol. J.*, **11**, 211–222. <https://doi.org/10.1111/pbi.12025>
- Haley, C. (2011) A cornucopia of maize genes. *Nat. Genet.*, **43**, 87–88. <https://doi.org/10.1038/ng0211-87>
- Hall, S.D., Stanford, I.M., Yamawaki, N., McAllister, C.J., Rönnqvist, K.C., Woodhall, G.L. and Furlong, P.L. (2011) The role of GABAergic modulation in motor function related neuronal network activity. *NeuroImage*, **56**, 1506–1510. <https://doi.org/10.1016/j.neuroimage.2011.02.025>
- Henry, C., Bledsoe, S.W., Siekman, A., Kollman, A., Waters, B.M., Feil, R., Stitt, M. and Lagrimini, L.M. (2014) The trehalose pathway in maize: Conservation and gene regulation in response to the diurnal cycle and extended darkness. *J. Exp. Bot.*, **65**, 5959–5973. <https://doi.org/10.1093/jxb/eru335>
- 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>
- Holding, D.R., Meeley, R.B., Hazebroek, J., Selinger, D., Gruis, F., Jung, R. and Larkins, B.A. (2010) Identification and characterization of the maize arogenate dehydrogenase gene family. *J. Exp. Bot.*, **61**, 3663–3673.
- Jander, G. and Joshi, V. (2009) Aspartate-Derived Amino Acid Biosynthesis in Arabidopsis thaliana. *Arabidopsis Book*, **7**, e0121. <https://doi.org/10.1199/tab.0121>
- Kemper, E.L., Neto, G.C., Papes, F., Moraes, K.C., Leite, A. and Arruda, P. (1999) The role of opaque2 in the control of lysine-degrading activities in developing maize endosperm. *Plant Cell*, **11**, 1981–1994. <https://doi.org/10.1105/tpc.11.10.1981>
- Keurentjes, J.J. (2009) Genetical metabolomics: Closing in on phenotypes. *Curr. Opin. Plant Biol.*, **12**, 223–230. <https://doi.org/10.1016/j.pbi.2008.12.003>
- 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>
- Li, M.X., Yeung, J.M., Cherny, S.S. and Sham, P.C. (2012) Evaluating the effective numbers of independent tests and significant *P*-value thresholds in commercial genotyping arrays and public imputation reference datasets. *Hum. Genet.*, **131**, 747–756. <https://doi.org/10.1007/s00439-011-1118-2>
- 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>
- Lisec, J., Meyer, R.C., Steinfath, M. *et al.* (2008) Identification of metabolic and biomass QTL in Arabidopsis thaliana in a parallel analysis of RIL and IL populations. *Plant J.*, **53**, 960–972.
- Liu, H., Luo, X., Niu, L. *et al.* (2017) Distant eQTLs and Non-coding Sequences Play Critical Roles in Regulating Gene Expression and Quantitative Trait Variation in Maize. *Mol. Plant*, **10**, 414–426. <https://doi.org/10.1016/j.molp.2016.06.016>
- Livak, K.J. and Schmittgen, T.D. (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods*, **25**, 402–408. <https://doi.org/10.1006/meth.2001.1262>
- Lunn, J.E., Delorge, I., Figueroa, C.M., van Dijk, P. and Stitt, M. (2014) Trehalose metabolism in plants. *Plant J.*, **79**, 544–567. <https://doi.org/10.1111/tpj.12509>
- Maeda, H. and Dudareva, N. (2012) The shikimate pathway and aromatic amino acid biosynthesis in plants. *Annu. Rev. Plant Biol.*, **63**, 73–105. <https://doi.org/10.1146/annurev-arplant.042811-105439>
- Maeda, H., Shasany, A.K., Schnepf, J., Orlova, I., Taguchi, G., Cooper, B.R., Rhodes, D., Pichersky, E. and Dudareva, N. (2010) RNAi suppression of Arogenate Dehydratase1 reveals that phenylalanine is synthesized predominantly via the arogenate pathway in petunia petals. *Plant Cell*, **22**, 832–849. <https://doi.org/10.1105/tpc.109.073247>
- Mertz, E.T., Bates, L.S. and Nelson, O.E. (1964) Mutant gene that changes protein composition and increases lysine content of maize endosperm. *Science*, **145**, 279–280. <https://doi.org/10.1126/science.145.3629.279>
- Nuccio, M.L., Wu, J., Mowers, R. *et al.* (2015) Expression of trehalose-6-phosphate phosphatase in maize ears improves yield in well-watered and drought conditions. *Nat. Biotechnol.*, **33**, 862–869. <https://doi.org/10.1038/nbt.3277>
- Pick, T.R., Brautigam, A., Schluter, U. *et al.* (2011) Systems analysis of a maize leaf developmental gradient redefines the current C4 model and provides candidates for regulation. *Plant Cell*, **23**, 4208–4220. <https://doi.org/10.1105/tpc.111.090324>
- Reyes, A.R., Bonin, C.P., Houmar, N.M., Huang, S. and Malvar, T.M. (2009) Genetic manipulation of lysine catabolism in maize kernels. *Plant Mol. Biol.*, **69**, 81–89. <https://doi.org/10.1007/s11103-008-9409-2>
- Riedelshimer, C., Lisec, J., Czedik-Eysenberg, A., Sulpice, R., Flis, A., Grieder, C., Altmann, T., Stitt, M., Willmitzer, L. and Melchinger, A.E. (2012) Genome-wide association mapping of leaf metabolic profiles for dissecting complex traits in maize. *Proc. Natl. Acad. Sci. U. S. A.*, **109**, 8872–8877. <https://doi.org/10.1073/pnas.1120813109>
- Roessner, U., Luedemann, A., Brust, D., Fiehn, O., Linke, T., Willmitzer, L. and Fernie, A. (2001) Metabolic profiling allows comprehensive phenotyping of genetically or environmentally modified plant systems. *Plant Cell*, **13**, 11–29. <https://doi.org/10.1105/tpc.13.1.11>
- Saito, K. and Matsuda, F. (2010) Metabolomics for functional genomics, systems biology, and biotechnology. *Annu. Rev. Plant Biol.*, **61**, 463–489. <https://doi.org/10.1146/annurev-arplant.043008.092035>
- Sauvage, C., Segura, V., Bauchet, G., Stevens, R., Do, P.T., Nikoloski, Z., Fernie, A.R. and Causse, M. (2014) Genome-Wide Association in Tomato Reveals 44 Candidate Loci for Fruit Metabolic Traits. *Plant Physiol.*, **165**, 1120–1132. <https://doi.org/10.1104/pp.114.241521>
- 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>
- Schauer, N., Semel, Y., Balbo, I., Steinfath, M., Reipsilber, D., Selbig, J., Pleban, T., Zamir, D. and Fernie, A.R. (2008) Mode of inheritance of primary metabolic traits in tomato. *Plant Cell*, **20**, 509–523. <https://doi.org/10.1105/tpc.107.056523>
- Schenck, C.A., Holland, C.K., Schneider, M.R., Men, Y., Lee, S.G., Jez, J.M. and Maeda, H.A. (2017) Molecular basis of the evolution of alternative tyrosine biosynthetic routes in plants. *Nat. Chem. Biol.*, **13**, 1029–1035. <https://doi.org/10.1038/nchembio.2414>

- Schluepmann, H. and Paul, M. (2009) Trehalose Metabolites in Arabidopsis-elusive, active and central. *Arabidopsis Book*, **7**, e0122. <https://doi.org/10.1199/tab.0122>
- Schmidt, R.J., Ketudat, M., Aukerman, M.J. and Hoschek, G. (1992) Opaque-2 is a transcriptional activator that recognizes a specific target site in 22-kD zein genes. *Plant Cell*, **4**, 689–700. <https://doi.org/10.1105/tpc.4.6.689>
- Sekhon, R.S., Lin, H., Childs, K.L., Hansey, C.N., Buell, C.R., de Leon, N. and Kaeppler, S.M. (2011) Genome-wide atlas of transcription during maize development. *Plant J.*, **66**, 553–563. <https://doi.org/10.1111/j.1365-313X.2011.04527.x>
- Shi, S.Q., Shi, Z., Jiang, Z.P., Qi, L.W., Sun, X.M., Li, C.X., Liu, J.F., Xiao, W.F. and Zhang, S.G. (2010) Effects of exogenous GABA on gene expression of Caragana intermedia roots under NaCl stress: Regulatory roles for H₂O₂ and ethylene production. *Plant, Cell Environ.*, **33**, 149–162. <https://doi.org/10.1111/j.1365-3040.2009.02065.x>
- Stepansky, A., Less, H., Angelovici, R., Aharon, R., Zhu, X. and Galili, G. (2006) Lysine catabolism, an effective versatile regulator of lysine level in plants. *Amino Acids*, **30**, 121–125. <https://doi.org/10.1007/s00726-005-0246-1>
- Takayama, M. and Ezura, H. (2015) How and why does tomato accumulate a large amount of GABA in the fruit? *Front Plant Sci.*, **6**, 612.
- Takayama, M., Matsukura, C., Ariizumi, T. and Ezura, H. (2017) Activating glutamate decarboxylase activity by removing the autoinhibitory domain leads to hyper γ -aminobutyric acid (GABA) accumulation in tomato fruit. *Plant Cell Rep.*, **36**, 103–116. <https://doi.org/10.1007/s00299-016-2061-4>
- Tieman, D., Zhu, G., Resende, M.F. Jr et al. (2017) A chemical genetic roadmap to improved tomato flavor. *Science*, **355**, 391–394. <https://doi.org/10.1126/science.aal1556>
- Tzin, V. and Galili, G. (2010) New insights into the shikimate and aromatic amino acids biosynthesis pathways in plants. *Mol. Plant*, **3**, 956–972. <https://doi.org/10.1093/mp/ssq048>
- Walley, J.W., Sartor, R.C., Shen, Z. et al. (2016) Integration of omic networks in a developmental atlas of maize. *Science*, **353**, 814–818. <https://doi.org/10.1126/science.aag1125>
- Wang, L., Czedik-Eysenberg, A., Mertz, R.A. et al. (2014) Comparative analyses of C₄ and C₃ photosynthesis in developing leaves of maize and rice. *Nat. Biotechnol.*, **32**, 1158–1165. <https://doi.org/10.1038/nbt.3019>
- 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.
- 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., Brotman, Y., Willmitzer, L., Yan, J. and Fernie, A.R. (2016a) Broadening Our Portfolio in the Genetic Improvement of Maize Chemical Composition. *Trends Genet.*, **32**, 459–469. <https://doi.org/10.1016/j.tig.2016.05.003>
- Wen, W., Liu, H., Zhou, Y. et al. (2016b) Combining Quantitative Genetics Approaches with Regulatory Network Analysis to Dissect the Complex Metabolism of the Maize Kernel. *Plant Physiol.*, **170**, 136–146. <https://doi.org/10.1104/pp.15.01444>
- Xiao, Y., Liu, H., Wu, L., Warburton, M. and Yan, J. (2017) Genome-wide Association Studies in Maize: Praise and Stargaze. *Mol. Plant*, **10**, 359–374. <https://doi.org/10.1016/j.molp.2016.12.008>
- Yang, X., Gao, S., Xu, S., Zhang, Z., Prasanna, B.M., Li, L., Li, J. and Yan, J. (2010) Characterization of a global germplasm collection and its potential utilization for analysis of complex quantitative traits in maize. *Mol. Breeding*, **28**, 511–526.
- Yang, Q.Q., Zhang, C.Q., Chan, M.L. et al. (2016) Biofortification of rice with the essential amino acid lysine: Molecular characterization, nutritional evaluation, and field performance. *J. Exp. Bot.*, **67**, 4285–4296. <https://doi.org/10.1093/jxb/erw209>
- Yu, G.H., Zou, J., Feng, J., Peng, X.B., Wu, J.Y., Wu, Y.L., Palanivelu, R. and Sun, M.X. (2014) Exogenous gamma-aminobutyric acid (GABA) affects pollen tube growth via modulating putative Ca²⁺-permeable membrane channels and is coupled to negative regulation on glutamate decarboxylase. *J. Exp. Bot.*, **65**, 3235–3248. <https://doi.org/10.1093/jxb/eru171>
- Zhang, Z., Ersoz, E., Lai, C.Q. et al. (2010) Mixed linear model approach adapted for genome-wide association studies. *Nat. Genet.*, **42**, 355–360. <https://doi.org/10.1038/ng.546>
- 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>
- Zhou, M.L., Zhang, Q., Sun, Z.M., Chen, L.H., Liu, B.X., Zhang, K.X., Zhu, X.M., Shao, J.R., Tang, Y.X. and Wu, Y.M. (2013) Trehalose Metabolism-Related Genes in Maize. *J. Plant Growth Regul.*, **33**, 256–271.
- Zhu, F.Y., Chen, M.X., Ye, N.H. et al. (2017) Proteogenomic analysis reveals alternative splicing and translation as part of the abscisic acid response in Arabidopsis seedlings. *Plant J.*, **91**(3), 518–533. <https://doi.org/10.1111/tbj.13571>