The Plant Cell, Vol. 31: 1990–2009, September 2019, www.plantcell.org © 2019 ASPB.

LARGE-SCALE BIOLOGY ARTICLE

Evolutionary Metabolomics Identifies Substantial Metabolic Divergence between Maize and Its Wild Ancestor, Teosinte[OPEN]

Guanghui Xu,[a,1] Jingjing Cao,[a,b,1] Xufeng Wang,[a] Qiuyue Chen,[a] Weiwei Jin,[a] Zhen Li,[a,2] and Feng Tian[a,2]

[a] State Key Laboratory of Plant Physiology and Biochemistry, National Maize Improvement Center, Key Laboratory of Biology and Genetic Improvement of Maize (MOA), Beijing Key Laboratory of Crop Genetic Improvement, Center for Crop Functional Genomics and Molecular Breeding, China Agricultural University, Beijing 100193, China
[b] Institute of Plant Protection, Jiangsu Academy of Agricultural Sciences, Nanjing 210014, China

ORCID IDs: 0000-0001-7986-7077 (G.X.); 0000-0001-5963-5041 (J.C.); 0000-0001-9345-827X (X.W.); 0000-0002-3304-8321 (Q.C.); 0000-0001-9320-9628 (W.J.); 0000-0002-4798-4161 (Z.L.); 0000-0003-3552-4536 (F.T.)

Maize (*Zea mays* subsp *mays*) was domesticated from its wild ancestor, teosinte (*Zea mays* subsp *parviglumis*). Maize's distinct morphology and adaptation to diverse environments required coordinated changes in various metabolic pathways. However, how the metabolome was reshaped since domestication remains poorly understood. Here, we report a comprehensive assessment of divergence in the seedling metabolome between maize and teosinte. In total, 461 metabolites exhibited significant divergence due to selection. Interestingly, teosinte and tropical and temperate maize, representing major stages of maize evolution, targeted distinct sets of metabolites. Alkaloids, terpenoids, and lipids were specifically targeted in the divergence between teosinte and tropical maize, while benzoxazinoids were specifically targeted in the divergence between tropical and temperate maize. To identify genetic factors controlling metabolic divergence, we assayed the seedling metabolome of a large maize-by-teosinte cross population. We show that the recent metabolic divergence between tropical and temperate maize tended to have simpler genetic architecture than the divergence between teosinte and tropical maize. Through integrating transcriptome data, we identified candidate genes contributing to metabolic divergence, many of which were under selection at the nucleotide and transcript levels. Through overexpression or mutant analysis, we verified the roles of *Flavanone 3-hydroxylase1*, *Purple aleurone1*, and maize *terpene synthase1* in the divergence of their related biosynthesis pathways. Our findings not only provide important insights into domestication-associated changes in the metabolism but also highlight the power of combining omics data for trait dissection.

INTRODUCTION

Plants produce structurally and functionally diverse metabolites that are important for plant growth, development, and adaptation to environmental changes (Dixon and Strack, 2003). These metabolites include primary metabolites that are essential for plant growth and reproduction and specialized metabolites that are closely associated with biotic and abiotic stress responses (Verpoorte and Memelink, 2002; Dixon and Strack, 2003; Schwab, 2003).

Uncovering the natural variation in plant metabolism and exploring the underlying genetic basis have received wide interest (Keurentjes, 2009; Fernie and Tohge, 2017; Fang et al., 2019). Using various single parental or multiparental linkage populations, many metabolite quantitative trait loci (mQTLs) in plant primary and secondary metabolism have been identified (Keurentjes et al., 2006; Schauer et al., 2006; Matsuda et al., 2012; Toubiana et al., 2012; Alseekh et al., 2013; Gong et al., 2013; Hill et al., 2015; Wen et al., 2015; Knoch et al., 2017; Rambla et al., 2017). These mQTL studies have revealed general features of the genetic architecture of plant metabolomes. However, due to the relatively low mapping resolution of linkage populations, the genes underlying mQTLs remain elusive. With significant advances in high-throughput genotyping technologies, metabolite-based genome-wide association studies in natural populations have become a powerful approach to dissect the genetic and biochemical bases of plant metabolomes (Chan et al., 2011; Riedelsheimer et al., 2012; Angelovici et al., 2013; Li et al., 2013; Chen et al., 2014, 2016; Sauvage et al., 2014; Wen et al., 2014; Matsuda et al., 2015; Tieman et al., 2017; Wu et al., 2018). Recent integration of metabolomics with other omics data has proven to be highly effective in functional gene identification and pathway elucidation (Wen et al., 2015; Zhu et al., 2018). These metabolite-based genome-wide association studies and integrative studies have identified novel genes for important metabolic traits and uncovered potential metabolic networks. Importantly, many metabolite features are closely associated with traits of agronomic importance and thus could be used as biomarkers in metabolomics-assisted

[1] These authors contributed equally to this work.
[2] Address correspondence to: ft55@cau.edu.cn and lizhenchem@cau.edu.cn.
The author responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors (www.plantcell.org) is: Feng Tian (ft55@cau.edu.cn).
[OPEN]Articles can be viewed without a subscription.
www.plantcell.org/cgi/doi/10.1105/tpc.19.00111

IN A NUTSHELL

Background: Studies of domestication have focused almost exclusively on changes in morphology such as increased grain size, loss of shattering, larger inflorescences with more grain, and reduced branching. Yet, domestication certainly involved a much larger set of traits that have escaped analysis because they are not visible phenotypes. Among the traits that have gone largely unanalyzed are changes in the metabolic profiles of crops as compared to their progenitors. In this study, we sought to fill this void by analyzing metabolic evolution between maize and its progenitor, teosinte, which is the best-studied crop-progenitor system in the literature.

Question: In this study, we aimed to identify the metabolites that were targeted by selection during maize domestication and elucidate the underlying genetic basis. We achieved this goal by measuring metabolite abundance in the seedlings from diverse maize and teosinte accessions as well as from a maize-by-teosinte cross population.

Findings: We showed that a total of 461 metabolites exhibited significant divergence due to selection. Teosinte, tropical, and temperate maize, representing major stages of maize evolution, targeted distinct sets of metabolites. Alkaloids, terpenoids, and lipids were specifically targeted in the divergence between teosinte and tropical maize, while benzoxazinoids were specifically targeted in the divergence between tropical and temperate maize. Through integrating genome, transcriptome, and metabolome data from a maize-by-teosinte cross population, we identified candidate genes contributing to metabolic divergence, many of which were under selection at the nucleotide and transcript levels. Through overexpression or mutant analysis, we verified the roles of *FHT1*, *Pr1*, and *ZmTPS1* in the divergence of their related biosynthesis pathways.

Next steps: Our evolutionary metabolomics study showed that the metabolome was reshaped during maize domestication and post-domestication adaptation. Further functional studies, such as experiments demonstrating the role of metabolites in biotic or abiotic resistance, are required to make direct connections between metabolite divergence and their ecological functions.

breeding (Riedelsheimer et al., 2012; Wen et al., 2014; Chen et al., 2016; Westhues et al., 2017).

Maize (*Zea mays* subsp *mays*) was domesticated in southwestern Mexico ~9000 years ago from its wild ancestor, teosinte (*Zea mays* subsp *parviglumis*; Matsuoka et al., 2002; Piperno et al., 2009). Compared with teosinte, maize underwent a dramatic morphological transformation in plant and inflorescence architecture during domestication (Doebley, 2004). In addition to the dramatic morphological changes, maize also experienced substantial changes in adaptation. From its tropical origin, maize has adapted to diverse environments and become the most productive crop worldwide (Hung et al., 2012; Guo et al., 2018; Huang et al., 2018). Several key genes controlling the major morphological changes and maize adaptation have been cloned (Wang et al., 2005; Studer et al., 2011; Hung et al., 2012; Wills et al., 2013; Yang et al., 2013; Guo et al., 2018; Huang et al., 2018; Liang et al., 2019.

Increasing evidence has shown that the significant phenotypic changes during crop domestication were frequently accompanied by coordinated metabolite alterations (Shang et al., 2014; Beleggia et al., 2016; Zhou et al., 2016; Zhu et al., 2018). For example, selection for nonbitter taste was associated with significant changes in steroidal glycoalkaloid and cucurbitacin compounds during the domestication of tomato (*Solanum lycopersicum*) and cucumber (*Cucumis sativus*), respectively (Shang et al., 2014; Zhou et al., 2016; Tieman et al., 2017; Zhu et al., 2018). Systematic studies have been performed to examine the impact of maize domestication on its transcriptome (Swanson-Wagner et al., 2012; Lemmon et al., 2014; Wang et al., 2018). The maize transcriptome has experienced substantial alterations since domestication, with numerous genes targeted by selection for altered expression, many of which are closely associated with local adaptation (Swanson-Wagner et al., 2012; Lemmon et al., 2014; Wang et al., 2018). By contrast, an understanding of the metabolic changes that accompanied the maize domestication process and their associated significances is generally lacking.

In this study, we used an evolutionary metabolomics method to comprehensively assess metabolite differences in seedlings of teosinte, tropical maize, and temperate maize, which represent major stages of maize evolution. We showed that distinct sets of metabolites were targeted by selection during the evolution from teosinte to tropical maize to temperate maize. Using a large maize-by-teosinte cross population, we performed mQTL mapping to identify the genetic factors driving metabolic divergence. Through integrating with the previous transcriptome data (Wang et al., 2018), we identified candidate genes underlying mQTLs. Further combing with the previous population genomics data (Hufford et al., 2012; Swanson-Wagner et al., 2012; Lemmon et al., 2014; Wang et al., 2018) revealed that these candidate genes for mQTLs are more likely targeted by selection at the nucleotide and gene expression levels since domestication. Finally, we validated the functions of several important metabolic genes and revealed the associated evolutionary significances. Our findings not only provide important insights into how domestication reshaped the maize metabolome but also highlight the power of combining omics data for trait dissection.

RESULTS AND DISCUSSION

Teosinte and Maize Exhibited Significant Divergence in Their Metabolomes

To determine the difference in the metabolomes between maize and teosinte, the seedling metabolomes of a diverse panel of 43 teosinte and maize accessions (Supplemental Table 1) were

quantified using a nontargeted liquid chromatography–high-resolution mass spectrometry (LC-HRMS)–based metabolomics approach. The teosinte set included 16 accessions of *Z. m.* subsp *parviglumis*. The maize set consisted of 12 tropical inbred lines and 15 temperate inbred lines representing a broad genetic diversity of maize.

In total, 2378 nonredundant metabolites were detected, including 291 annotated metabolites (Supplemental Data Set 1). These annotated metabolites covered a broad range of metabolite classes, including 58 carboxylic acids, 34 flavonoids, 28 amino acids, 25 lipids, 21 terpenoids, 19 alkaloids, 12 benzoxazinoids, 12 peptides, 8 alcohol metabolites, 6 benzenoids, 5 phenolic glycosides, 4 amines, 2 vitamins, 2 pyrans, and 55 unclassified metabolites (Figure 1A; Supplemental Data Set 1). We performed a principal component analysis (PCA) for the teosinte and maize accessions with all metabolites. Interestingly, the 43 teosinte and maize accessions separated into three clusters (Figure 1B). Further examination showed that each cluster respectively contained teosinte, tropical maize, and temperate maize lines, indicating a significant metabolic divergence among them. Moreover, the tropical maize cluster resided between the teosinte and temperate maize clusters, indicating that a gradual metabolic divergence might have occurred during the evolution from teosinte to tropical maize to temperate maize. Similar results were obtained when PCA was performed with the 291 annotated metabolites (Supplemental Figure 1). We referred to the metabolic divergence between teosinte and tropical maize as Teosinte-TroMaize and the metabolic divergence between tropical and temperate maize as TroMaize-TemMaize. These two divergences arose subsequently and represent two major stages of maize evolution. Given these overall features in metabolic divergence, our subsequent analysis focused on the comparison of Teosinte-TroMaize and TroMaize-TemMaize.

Distinct Sets of Metabolites Were Targeted in the Divergence of Teosinte-TroMaize and TroMaize-TemMaize

To identify metabolites that are responsible for the metabolic divergence in Teosinte-TroMaize and TroMaize-TemMaize, we used a Q_{ST}-F_{ST} comparison strategy (see "Methods") to identify the specific metabolites that were targeted by selection in Teosinte-TroMaize and TroMaize-TemMaize. Q_{ST} estimates the phenotypic divergence between taxa for the metabolite of interest, while F_{ST} estimates the molecular divergence between taxa. The Q_{ST}-F_{ST} method determines whether the observed differences in metabolite levels between taxa were caused by selection or by neutral processes. To estimate the level of neutral population divergence, the 43 teosinte and maize accessions were genotyped with ~6000 single-nucleotide polymorphisms (SNPs) using the Maize6KSNP chip. Those SNPs that deviate from neutrality were excluded through permutations, and a neutral F_{ST} was computed using the remaining SNPs and used as a neutral benchmark in the comparison of Teosinte-TroMaize and TroMaize-TemMaize. To determine the significance of the observed Q_{ST} relative to F_{ST}, a 99% bootstrapping support interval was generated for each Q_{ST} value. Only those metabolites whose 99% bootstrapping support interval of Q_{ST} was greater than the neutral F_{ST} value were considered targets of selection.

Of the 2378 nonredundant metabolites, 461 (19.4%) metabolites showed significant differences in metabolite contents due to selection in the comparison of Teosinte-TroMaize and TroMaize-TemMaize (Figure 2A), including 198 metabolites differentiating teosinte and tropical maize and 296 metabolites differentiating tropical and temperate maize (Figure 2B). This result suggests that the metabolic divergence between tropical and temperate maize, which arose more recently than the metabolic divergence between teosinte and tropical maize, involved more metabolite alterations. Compared with teosinte and tropical maize that were distributed in similar tropical environments, tropical and temperate maize differed more significantly in growing environments. As maize spread into temperate zones, maize was faced with more diverse conditions in temperature, daylength, and disease susceptibility (Liu et al., 2015). The metabolite alterations between tropical and temperate maize might have helped drive maize to rapidly adapt to wide temperate environments. Interestingly, a recent evolutionary metabolomics study for tomato fruits revealed similar results, in which 614 metabolites were significantly altered in the improvement stage of tomato, whereas 389 metabolites exhibited significant divergence in the domestication stage (Zhu et al., 2018). However, Beleggia et al. (2016) observed an opposite pattern in

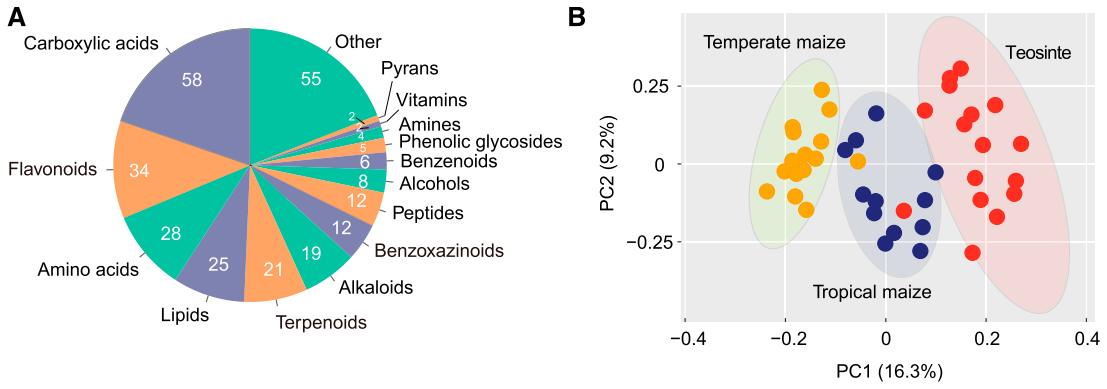

Figure 1. Metabolome Divergence between Maize and Teosinte.

(A) Classification of metabolites that have annotated structures. **(B)** PCA of the maize and teosinte accessions with all metabolites. PC, principal component.

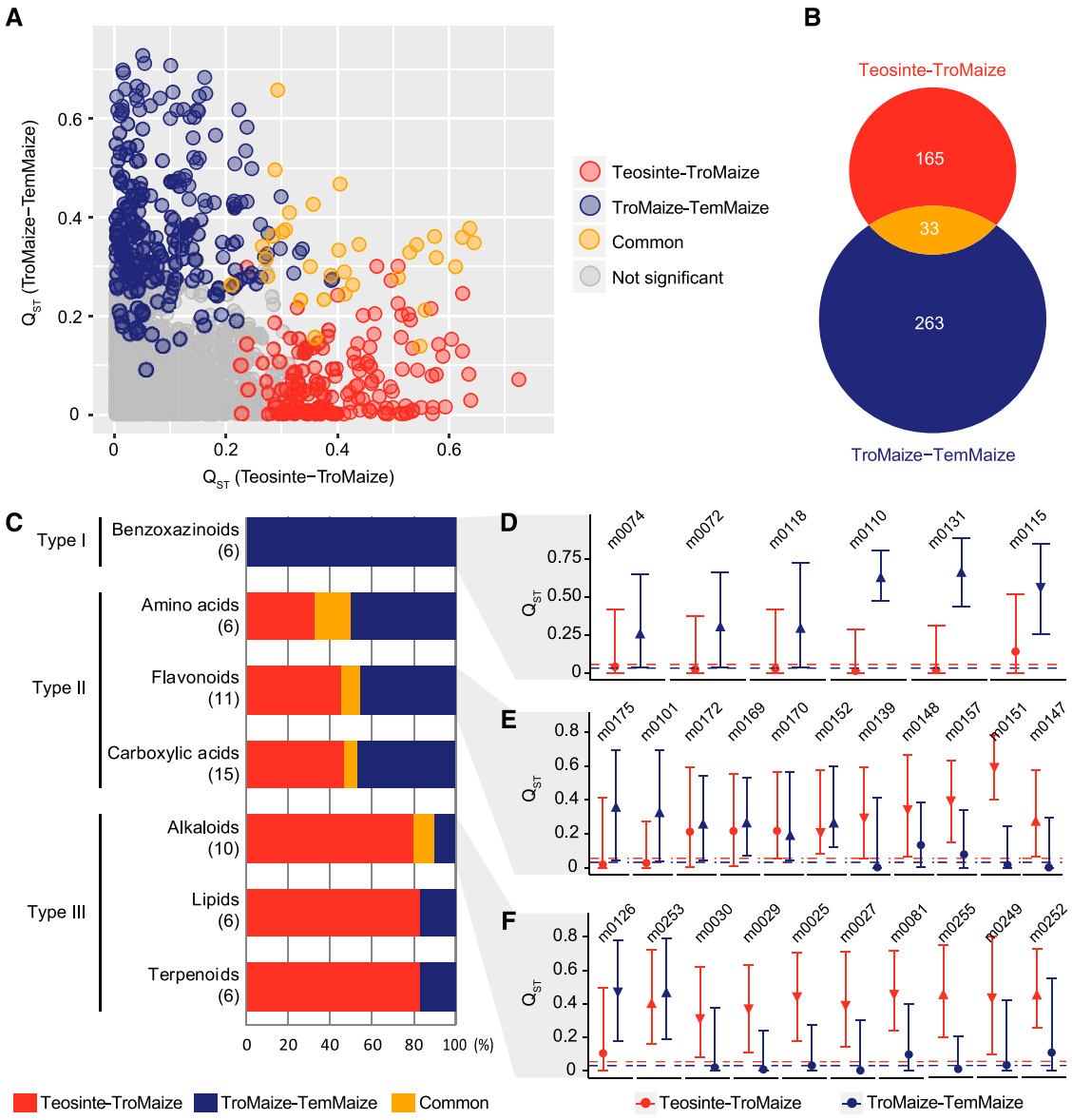

Figure 2. Identification and Characterization of Divergent Metabolites.

(A) Q_{ST} distribution associated with the two evolutionary steps: Teosinte-TroMaize and TroMaize-TemMaize. For each metabolite, Q_{ST} was calculated separately for Teosinte-TroMaize and TroMaize-TemMaize. The metabolites that exhibited divergence in Teosinte-TroMaize, TroMaize-TemMaize, and both processes are indicated in red, blue, and orange, respectively.
(B) Venn diagram showing the number of divergent metabolites in the two evolutionary processes.
(C) Different metabolite classes exhibited distinct divergence patterns. The bar chart shows the proportion of divergent metabolites in the corresponding evolutionary process. Numbers in parentheses are the number of annotated compounds in the corresponding metabolite class.
(D) to **(F)** Illustration of the divergent pattern using benzoxazinoids **(D)**, flavonoids **(E)**, and alkaloids **(F)** as examples. Bars represent 99% CIs of Q_{ST}. Dots and triangles on the bars represent observed Q_{ST}. Red dots indicate no significant differences between Q_{ST} and F_{ST} in Teosinte-TroMaize, and blue dots indicate no significant differences between Q_{ST} and F_{ST} in TroMaize-TemMaize. In Teosinte-TroMaize, upward and downward red triangles indicate significantly higher and lower metabolite content in tropical maize relative to teosinte, respectively. In TroMaize-TemMaize, upward and downward blue triangles indicate significantly higher and lower metabolite content in temperate maize relative to tropical maize, respectively. The horizontal red and blue dashed lines indicate the genome-wide threshold of neutral F_{ST} for Teosinte-TroMaize and TroMaize-TemMaize, respectively.

the analysis of the changes of kernel primary metabolites during the domestication of tetraploid wheat (*Triticum turgidum*), with 16 metabolites targeted by selection during domestication of emmer (primary domestication) but only 6 metabolites targeted during domestication of durum wheat (secondary domestication). One important difference among these three studies is the major types of metabolites captured. Zhu et al. (2018) and our study used LC-mass spectrometry (MS) for metabolite profiling, which mainly captured secondary metabolites, whereas Beleggia et al. (2016) used gas chromatography-MS which mainly captured primary

metabolites. It is well known that primary metabolites are essential for the survival of the plant, whereas secondary metabolites are closely associated with environmental responses. Therefore, it is highly likely that primary metabolites tend to be modified in the earlier stages of crop domestication, while secondary metabolites are more likely to be modified as crops spread into new environments. The different evolutionary patterns observed in these three studies might reflect the distinct evolutionary history associated with primary and secondary metabolism, but this hypothesis needs further testing by simultaneously profiling primary and secondary metabolites in cultivated crops and their wild ancestors, which will also provide a more complete picture of metabolome divergence since crop domestication.

Among the 461 metabolites that showed evidence of selection, only 33 (7.2%) metabolites were common targets of selection in both Teosinte-TroMaize and TroMaize-TemMaize (Figures 2A and 2B), indicating that different sets of metabolites were targeted during the evolutionary process of Teosinte-TroMaize and TroMaize-TemMaize. To determine which class of metabolites was involved in each divergent process, we analyzed the distribution of the 76 divergent metabolites that have annotated structures. The 76 annotated metabolites can be classified into 10 metabolite classes (Supplemental Data Set 1). Our subsequent analysis at the metabolite class level mainly focused on the seven metabolite classes that contain at least five divergent compounds, including benzoxazinoids, carboxylic acids, flavonoids, amino acids, alkaloids, terpenoids, and lipids (Figure 2C). Overall, these seven metabolite classes exhibited three types of evolutionary features (Figure 2C).

The first type (Type I) of metabolite class included benzoxazinoids that were specifically targeted in TroMaize-TemMaize. Of 12 annotated benzoxazinoid compounds, six compounds exhibited significant divergence between tropical and temperate maize but similar levels between teosinte and tropical maize (Figure 2D). This result suggested that the benzoxazinoid pathway might have experienced a specific divergence as maize spread from tropical regions to temperate zones. Notably, 2,4-dihydroxy-7-methoxy-1,4-benzoxazin-3-one glucoside (DIMBOA-Glc) and its upstream metabolites 2,4-dihydroxy-7-methoxy-1,4-benzoxazin-3-one (DIMBOA), 2,4-dihydroxy-7-methoxy-1,4-benzoxazin-one (DIBOA), 2,4,7-trihydroxy-2*H*-1,4-benzoxazin-3-(4*H*)-one (TRIBOA), and 2,4,7-trihydroxy-2*H*-1,4-benzoxazin-3-(4*H*)-one glucoside (TRIBOA-Glc) consistently exhibited higher contents in temperate maize compared with tropical maize, whereas 2-hydroxy-4,7-dimethoxy-1,4-benzoxazin-3-one (HDMBOA), converted from DIMBOA-Glc, showed higher content in tropical maize (Figure 2D). Accumulating more DIMBOA-Glc might be a local adaptation for temperate maize to resist insect herbivores in temperate environments.

The second type (Type II) of metabolite class included metabolite classes that were involved in both Teosinte-TroMaize and TroMaize-TemMaize, including carboxylic acids, flavonoids, and amino acids (Figure 2C). Within these metabolite classes, different metabolites were separately targeted in Teosinte-TroMaize and TroMaize-TemMaize. For example, 11 annotated metabolites in the flavonoid class exhibited significant divergence; among them, five metabolites were specifically targeted in Teosinte-TroMaize, five were specifically targeted in TroMaize-TemMaize, and one was commonly targeted in both evolutionary processes (Figure 2E). These divergent metabolites exhibited a complex divergence pattern in the directionality of the associated metabolic changes. Among the five flavonoid metabolites specifically targeted in Teosinte-TroMaize, four metabolites (m0139, tricin 7-diglucuronoside; m0148, kaempferol; m0151, quercetin derivative; and m0157, tricetin 3′-methyl ether 7-glucoside) exhibited higher contents in teosinte than in tropical maize, whereas the other metabolite (m0147, coumarin derivative) showed an opposite pattern (Figure 2E). All five metabolites that showed specific divergence in TroMaize-TemMaize [m0101, L-β-aspartyl-L-phenylalanine; m0169, puerarin 4′-*O*-glucoside; m0170, biochanin A 7-(6-methylmalonylglucoside); m0172, apigenin 4′-glucoside; and m0175, koparin] had higher contents in temperate maize than in tropical maize (Figure 2E). The metabolite (m0152, a chrysin derivative) that was commonly targeted in both evolutionary processes exhibited a decrease from teosinte to tropical maize but an increase from tropical maize to temperate maize (Figure 2E). These results suggested that the flavonoid pathway might have experienced complex modifications since domestication, in which different metabolic steps were targeted for local adaptation in different stages of maize evolution. It is interesting to note that two insect-deterrent flavonols (m0148, kaempferol and m0151, a quercetin derivative) and two UV-B–deterrent flavone *O*-glucosides (m0139, tricin 7-diglucuronoside and m0157, tricetin 3′-methyl ether 7-glucoside) exhibited decreased levels in tropical maize compared with teosinte (Figure 2E). These results suggested that teosinte is an important source to improve the biotic and abiotic resistance of maize.

The third type (Type III) included metabolite classes that were specifically targeted in Teosinte-TroMaize, including alkaloids, terpenoids, and lipids (Figure 2C). For example, 10 annotated metabolites in the alkaloid class exhibited significant divergence. Among them, eight were specifically targeted in Teosinte-TroMaize, only one metabolite was specifically targeted in TroMaize-TemMaize, and one was commonly targeted in both evolutionary processes (Figure 2F). Notably, all of these divergent alkaloid metabolites in Teosinte-TroMaize, except phosphocholine (m0252) and a uridine derivative (m0255), exhibited reduced contents in tropical maize compared with teosinte (Figure 2F). Many alkaloid compounds were associated with bitter taste (Chen et al., 2007; Frick et al., 2017). The consistently reduced alkaloid contents in maize compared with teosinte could be the result of selection for non-bitter grain, which consequently caused an associated reduction in seedling alkaloids. This reduced alkaloid content for improved palatability has been also observed during the domestication of potato (*Solanum tuberosum*; Johns and Alonso, 1990), lupins (Enneking and Wink, 2000), and tomato (Tieman et al., 2017; Zhu et al., 2018).

The Recent Metabolic Divergences between Tropical and Temperate Maize Were Controlled by Simpler Genetic Architecture

To identify the genetic factors controlling metabolite divergence, we further profiled the seedling metabolome of a maize-teosinte BC_2S_3 recombinant inbred line (RIL) population derived from a cross between the temperate maize inbred line W22 and the *Z. m*. subsp *parviglumis* accession CIMMYT 8759 (hereafter

referred to as 8759) using LC-HRMS (Supplemental Data Set 2). This RIL population consisted of 624 lines and was previously genotyped with 19,838 SNP markers using genotyping-by-sequencing technology (Shannon, 2012). The content of the 461 metabolites exhibiting significant divergence between maize and teosinte was determined for each RIL and used as the phenotypes to map mQTLs controlling the metabolite variation in the BC_2S_3 population. In total, 1494 mQTLs for 432 divergent metabolites were detected (Figure 3A; Supplemental Data Set 3).

We compared the genetic architecture underlying metabolites that exhibited specific divergence in the two evolutionary processes (Teosinte-TroMaize and TroMaize-TemMaize). Interestingly, the metabolites specifically targeted in TroMaize-TemMaize tended to be controlled by fewer mQTLs than the metabolites specifically targeted in Teosinte-TroMaize (Figure 3B; Student's *t* test, $P = 0.046$; permutation test, $P < 0.05$). Moreover, the mQTLs detected in TroMaize-TemMaize, on average, explained more genetic variance (proportion of phenotypic variance standardized by metabolite heritability) than the mQTLs detected in Teosinte-TroMaize ($P = 3.3E{-}6$, Figure 3C). These results suggested that the metabolic divergences in TroMaize-TemMaize, which arose more recently than the metabolic divergences in Teosinte-TroMaize, were controlled by simpler genetic architecture. Our observations further support the notion that phenotypic changes that came under strong and recent selection tend to have genetic

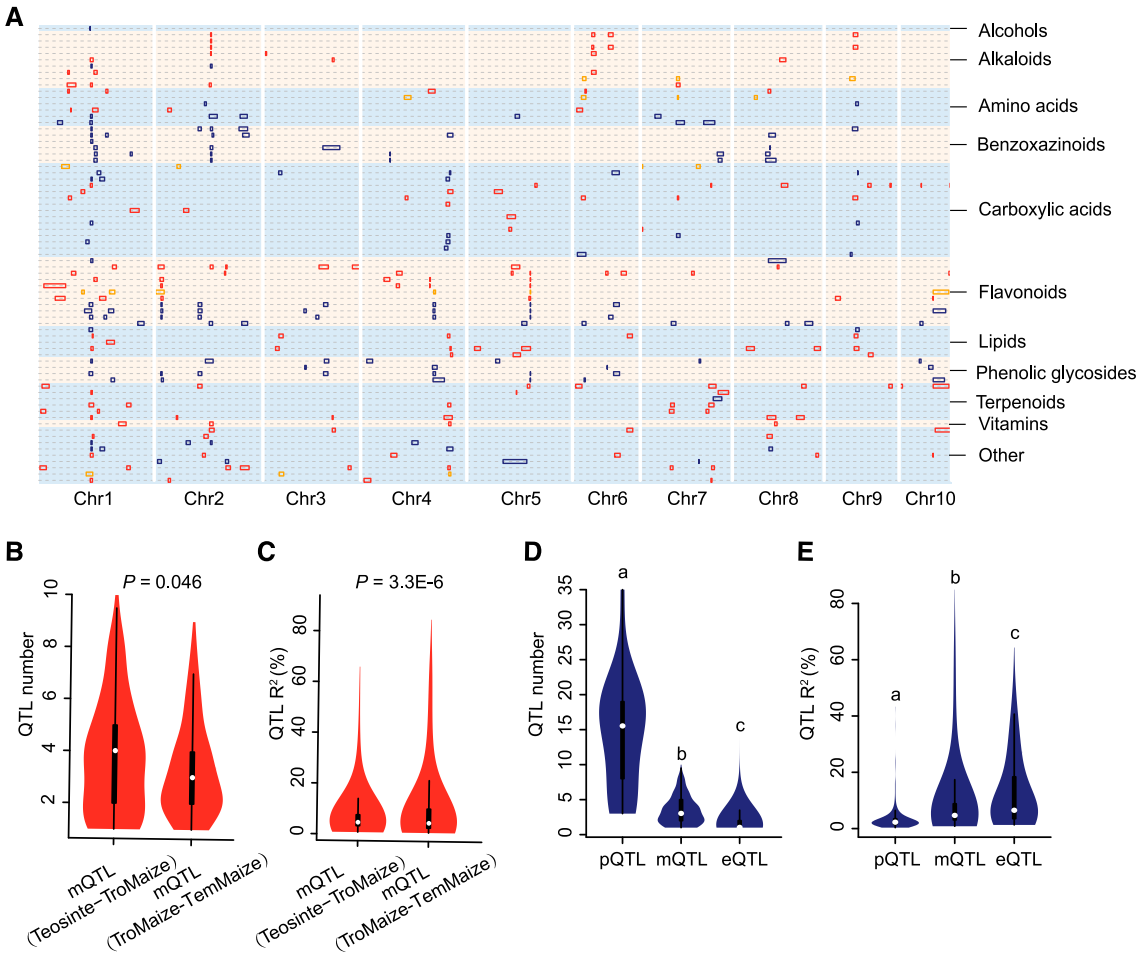

Figure 3. Genetic Architecture of Metabolite Divergence.

(A) mQTLs for divergent metabolites. Rectangles represent 2-LOD support intervals of the mQTLs. Divergent metabolites in Teosinte-TroMaize, TroMaize-TemMaize, and both processes are colored in red, blue, and orange, respectively. Only mQTLs for divergent metabolites that have annotated structures are shown here.

(B) and **(C)** Genetic architecture comparison for the metabolites specifically targeted in Teosinte-TroMaize and TroMaize-TemMaize. mQTL (Teosinte-TroMaize), mQTLs controlling the specific metabolic divergence in Teosinte-TroMaize; mQTL (TroMaize-TemMaize), mQTL controlling the specific metabolic divergence in TroMaize-TemMaize.

(D) and **(E)** Genetic architecture comparison for morphological phenotypes, metabolites, and gene expression. pQTLs, morphological phenotype QTLs reported in previous studies (Shannon, 2012; Huang et al., 2016; Li et al., 2016; Xu et al., 2017a, 2017b); mQTLs, QTLs for divergent metabolites identified in this study; eQTL, gene expression QTLs reported by Wang et al. (2018). Different letters indicate significant difference at $P < 0.01$ (Tukey test).

architectures with fewer genes and larger effect sizes than changes resulting from previous selection because of the shorter amount of time for the variation to accumulate (Wallace et al., 2014).

In the same maize-teosinte BC_2S_3 RIL population, we previously performed QTL mapping for 26 morphological phenotypes (termed pQTLs; Shannon, 2012; Huang et al., 2016; Li et al., 2016; Xu et al., 2017a, 2017b) and QTL mapping for 17,311 expressed genes (termed eQTLs; Wang et al., 2018). These data sets allowed us to compare the genetic architectures controlling natural variation in morphological phenotypes and intermediate molecular phenotypes such as metabolite and gene expression levels. On average, 14.2 pQTLs, 3.4 mQTLs, and 1.5 eQTLs were mapped for each trait of morphological phenotype, metabolite, and gene expression, respectively (Figure 3D). Each QTL averagely explained 3.2, 8.4, and 12.5% of the genetic variance of the trait of interest after accounting for trait heritability (Figure 3E). As expected, the morphological phenotype exhibited the highest complexity in genetic architecture, followed by metabolite and gene expression.

Analysis of the genomic distribution of the mQTLs showed that the mQTLs were not evenly distributed along the maize genome but formed 29 hotspots, each of which was associated with at least seven metabolites (Figure 4). Enrichment analysis for these mQTL hotspots revealed that the target metabolites regulated by an individual mQTL hotspot tended to be significantly enriched in specific metabolic pathways (Figure 4; Supplemental Table 2), indicating the presence of critical factors at the hotspot that have a broad impact on the enriched metabolic pathway. For example, the mQTL hotspots *hs2-1* and *hs5-3*, located on chromosome 2 and 5, respectively, exhibited significant enrichments in the flavonoid pathway (Figure 4). We peformed candidated gene analysis (next section) and found that *Flavanone 3-hydroxylase1* (*FHT1*) and *Purple aleurone1* (*Pr1*) are the most likely underlying genes for *hs2-1* and *hs5-3*, respectively.

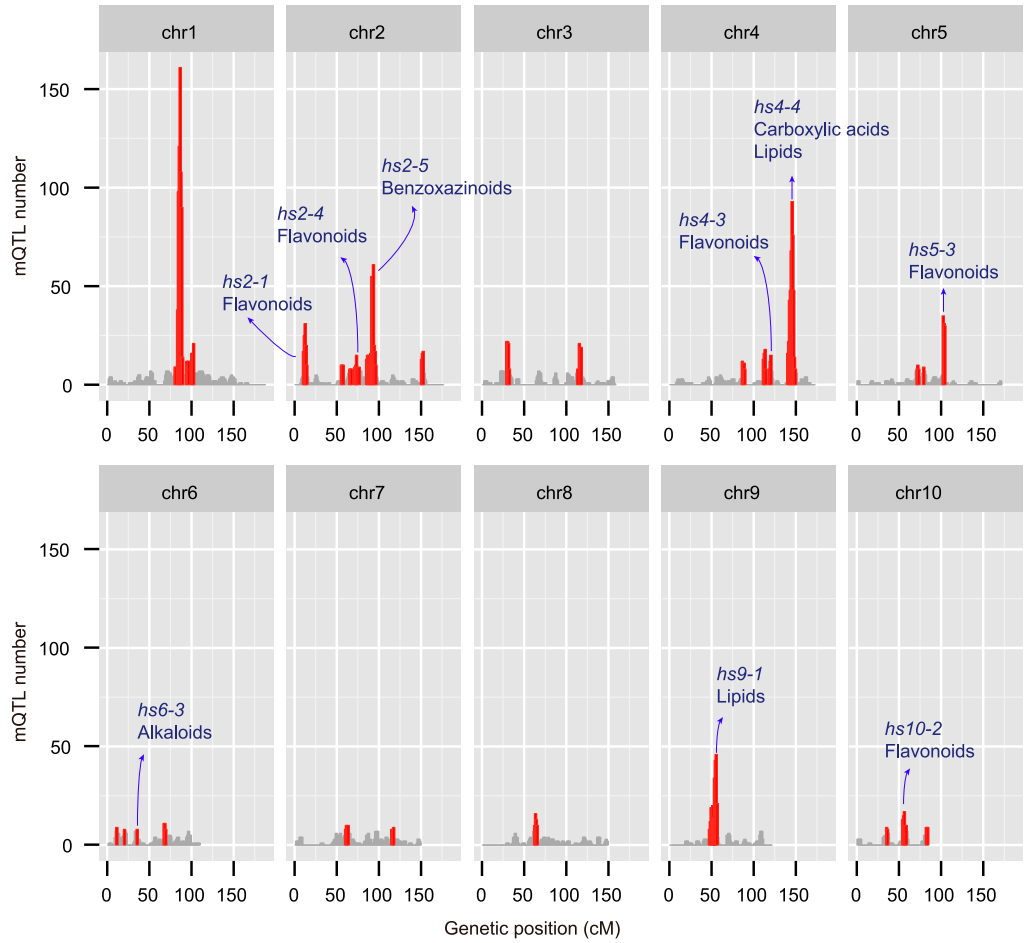

Figure 4. Distribution of mQTLs along the Maize Genome.

The *x* axis represents the genetic positions. We performed a sliding window analysis with 3-cM windows and 1-cM steps to count the number of mQTL in each window according to the peak position of the mQTLs. The *y* axis represents the number of mQTLs in each window, and windows with mQTL numbers greater or less than the mQTL hotspot threshold are in red and gray, respectively. Metabolic pathways enriched in the mQTL hotspots are in blue text.

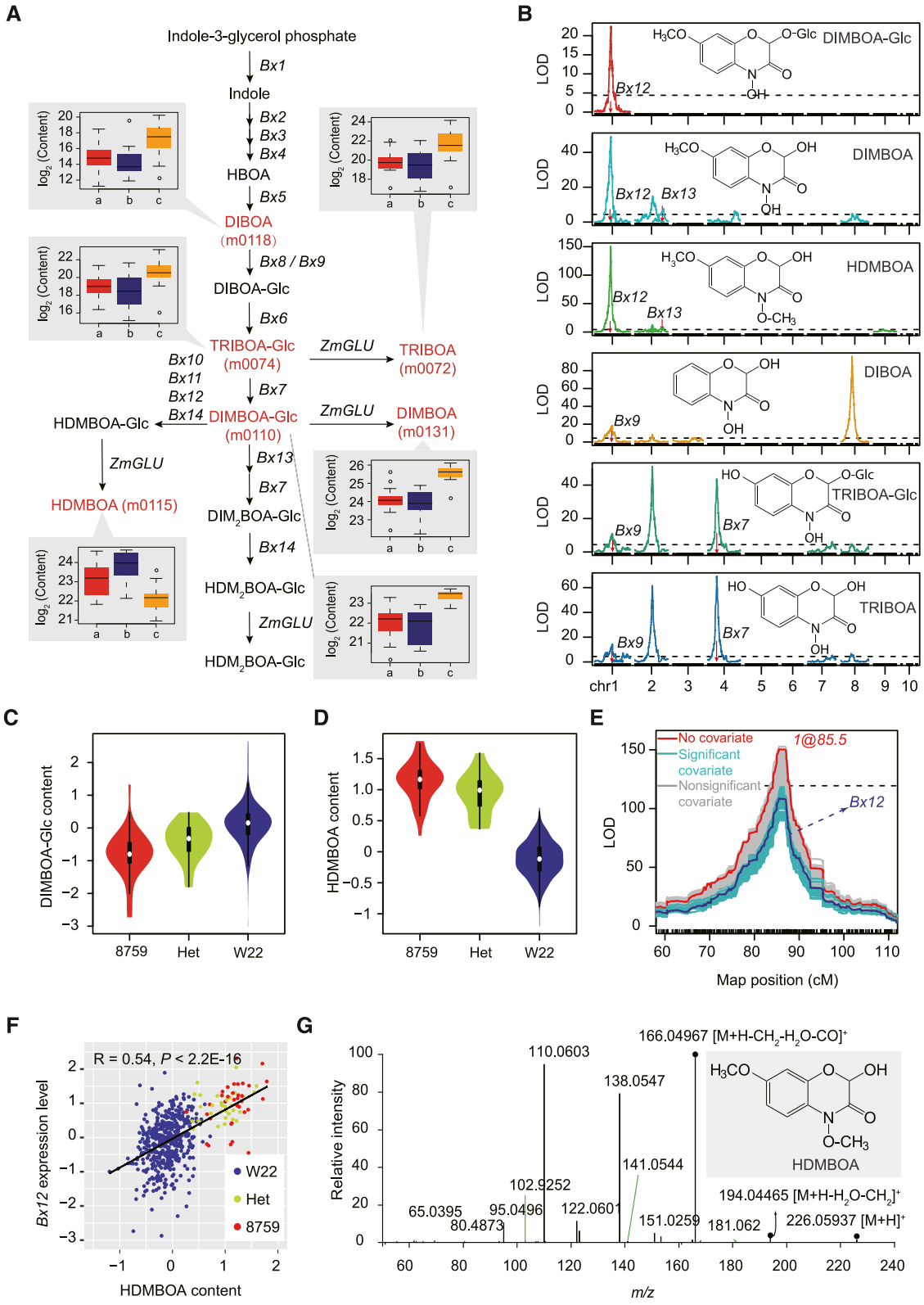

Figure 5. *Bx* Genes Contributed to the Metabolic Divergence in the Benzoxazinoid Pathway.

Integrating the Transcriptome to Identify Candidate Genes Controlling Metabolic Divergence

In our previous study, we performed transcriptome sequencing for the same seedling tissues used in this study for metabolic profiling (Wang et al., 2018), which enabled us to integrate transcriptome data to identify candidate genes underlying mQTLs. If the causal gene for an mQTL affects it through gene expression, the gene's expression level will significantly correlate with metabolite contents in the maize-teosinte BC_2S_3 RIL population. We used a previously proposed candidate gene ranking method (Lovell et al., 2015) to identify candidate genes that may underlie the target mQTL. The expression of each expressed gene within the 2-logarithm of odds (LOD) support interval of the target mQTL was iteratively added as a covariate to the original multiple-mQTL model to test the impact of gene expression on the identification of the target mQTL (Supplemental Figure 2). If there exists a causal link between gene expression and the target mQTL, adding gene expression to the mQTL model will significantly absorb the mQTL-specific phenotypic variance and consequently lead to a significant reduction in the LOD score of the mQTL (Supplemental Figure 2). The difference in mQTL LOD scores (designated $\triangle$LOD) between the original multiple-mQTL model and the new model adding a gene-expression covariate was used as a statistic to measure the relative effect of the candidate gene on the target mQTL (Supplemental Figure 2). To determine the significance of $\triangle$LOD, 10,000 permutations were performed for each mQTL.

Using this candidate gene ranking approach, we obtained ranked lists of candidate genes for 912 mQTLs (61% of mapped mQTLs; Supplemental Data Set 4). The number of candidate genes identified for each mQTL ranged from 1 to 23, with an average of 5.1 candidate genes per mQTL (Supplemental Data Set 4). Of the 912 mQTLs identified with candidate genes, 381 mQTLs (41.8%) were identified with fewer than three candidate genes, involving 409 gene models in total (Supplemental Data Set 4). To control the possible impact from unrelated background genes, our subsequent analysis of the entire set of candidate genes mainly focused on the 381 mQTLs identified with fewer than three candidate genes. A previous population genomics study identified a list of candidate genes that might be under selection at the nucleotide level during maize domestication and improvement (Hufford et al., 2012). Interestingly, of the 409 candidate genes, 45 (11%) were in the selection gene list (Supplemental Data Set 4), a finding that is unlikely to have occurred by chance alone (hypergeometric test, $P = 0.0084$). The previous comparative transcriptome analyses of maize and teosinte populations identified lists of genes that were potential selection targets for altered gene expression since domestication (Swanson-Wagner et al., 2012; Lemmon et al., 2014; Wang et al., 2018). We found that 149 (36.4%) candidate genes for mQTLs showed altered expression between maize and teosinte, a finding that is unexpected to occur by chance alone (hypergeometric test, $P = 2.29E-11$). These results strongly suggested that the candidate genes for mQTLs were more likely targeted by selection at both the nucleotide level and the expression level since domestication, which might have driven the metabolic divergence among teosinte, tropical maize, and temperate maize.

Bx Genes Contributed to the Metabolic Divergence in the Benzoxazinoid Pathway

Benzoxazinoids are specialized metabolites in grasses that function in defense against broad pests and pathogens (Niemeyer, 1988; Ahmad et al., 2011; Glauser et al., 2011; Meihls et al., 2013; Handrick et al., 2016). In maize seedlings, the predominant benzoxazinoid is DIMBOA, and it is stored as an inactive glucoside (DIMBOA-Glc) in the cell vacuole. In addition to DIMBOA-Glc, other benzoxazinoids, including HBOA-Glc, DIBOA-Glc, TRIBOA-Glc, DIM_2BOA-Glc, and HDMBOA-Glc, were also identified in maize leaves. DIMBOA-Glc often functions as a defensive signal that triggers the formation of callose to resist insect herbivores such as aphids (Ahmad et al., 2011; Meihls et al., 2013). HDMBOA-Glc is more toxic than DIMBOA-Glc and decays more quickly to produce insect-deterrent compounds against chewing herbivores such as caterpillars (Maresh et al., 2006; Ahmad et al., 2011; Meihls et al., 2013; Tzin et al., 2015). In our study, we found that six benzoxazinoid compounds exhibited specific divergence between tropical and temperate maize (Figure 2D). DIBOA, DIMBOA, DIMBOA-Glc, TRIBOA, and TRIBOA-Glc exhibited higher levels in temperate maize than in tropical maize, while HDMBOA showed higher levels in tropical maize (Figure 2D; Figure 5A).

Figure 5. (continued).

(A) to **(C)** Benzoxazinoid biosynthesis pathway and evidence of metabolite divergence. The boxplots next to the highlighted compounds denote the relative metabolite content in teosinte (a), tropical maize (b), and temperate maize (c).

(B) mQTLs mapped for the divergent benzoxazinoids.

(C) and **(D)** QTL effects of the common mQTL on chromosome 1 (*1@85.5*) for DIMBOA-Glc **(C)** and HDMBOA **(D)**. The red, blue, and green-yellow violin plots indicate RILs carrying homozygous 8759, homozygous W22, and heterozygous alleles at the mQTL *1@85.5* in the maize-teosinte BC_2S_3 population, respectively.

(E) Candidate gene analysis for the mQTL *1@85.5* for HDMBOA. The original LOD profile of *1@85.5* is plotted in red. The LOD profiles of *1@85.5* from new models adding gene expression covariate are plotted in turquoise (significant covariates) and gray (nonsignificant covariates). The dark blue line represents *Bx12*.

(F) Correlation between HDMBOA contents and *Bx12* expression levels in the maize-teosinte BC_2S_3 RIL population. RILs that are homozygous for 8759 and W22 and heterozygous at *Bx12* are indicated in red, blue, and green-yellow, respectively. The gene expression and metabolite content values are residuals after accounting for confounders.

(G) Structure and LC-MS/MS fragmentation of HDMBOA.

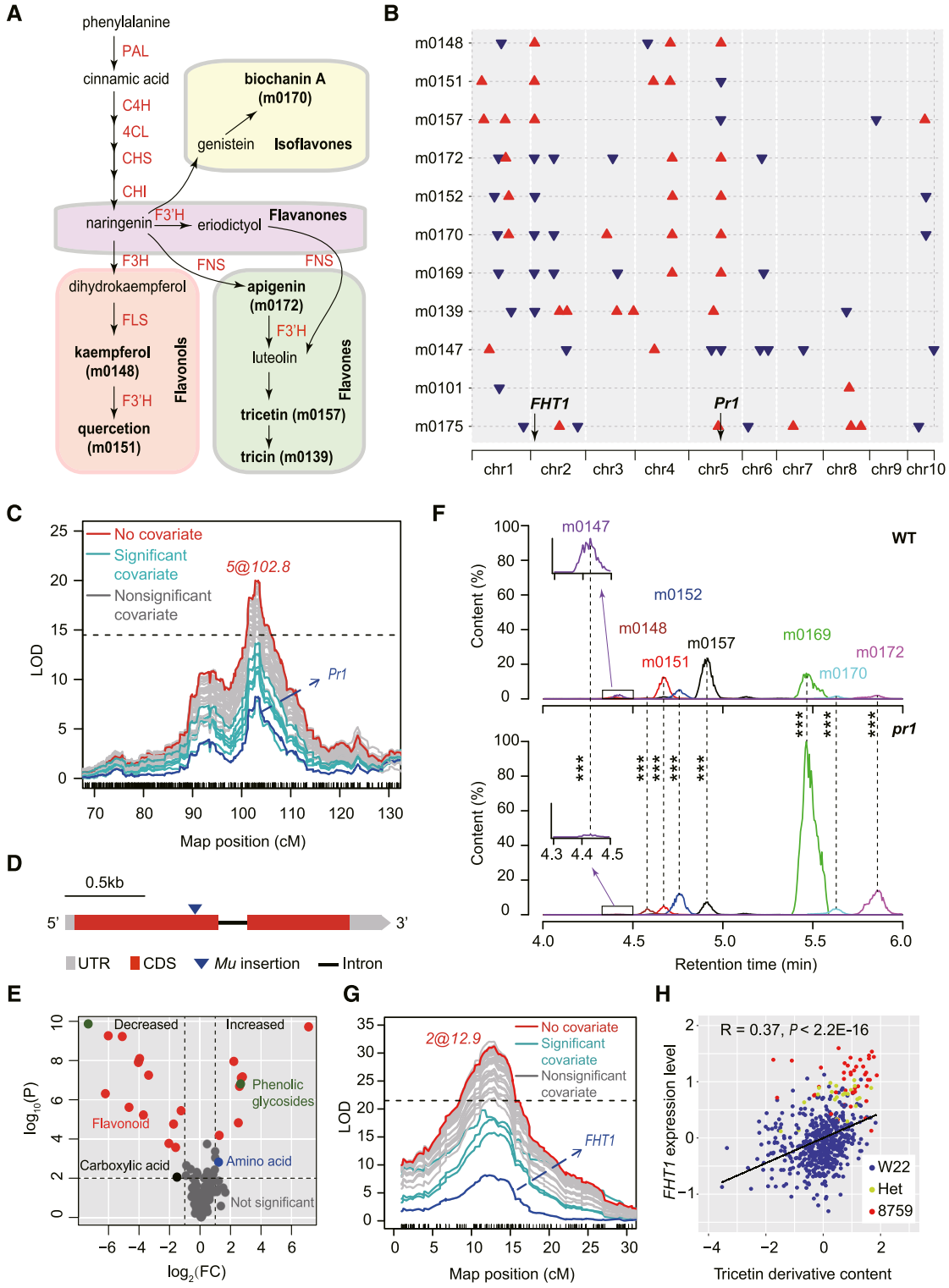

Figure 6. *FHT1* and *Pr1* Contributed to the Metabolic Divergence in the Flavonoid Pathway.

(A) Biosynthetic pathway for different structural forms of flavonoids. Divergent metabolites detected in this study are in bold. m0148, kaempferol; m0151, quercetin derivative; m0172, apigenin derivative; m0157, tricetin derivative; m0139, tricin derivative; m0170, biochanin A derivative. 4CL, 4-coumaryl-CoA ligase; C4H, cinnamic acid hydroxylase; CHI, chalcone isomerase; CHS, chalcone synthase; F3H, flavanone 3-hydroxylase; F3′H (PR1), flavanone 3′-hydroxylase; FNS, flavone synthase; PAL, phenylalanine-ammonia lyase.

To determine the genes controlling benzoxazinoid divergence, we performed mQTL mapping in the maize-teosinte BC_2S_3 population and identified 27 mQTLs for the six benzoxazinoid compounds (Figure 5B). Interestingly, three mQTLs for DIMBOA, DIMBOA-Glc, and HDMBOA colocalized in similar positions on chromosome 1 (Figure 5B). At this common mQTL (*1@85.5*), the W22 allele is associated with a higher content of DIMBOA-Glc but lower level of HDMBOA (Figures 5C and 5D). For each mQTL, putative candidate genes were identified using the candidate gene ranking approach. Interestingly, *benzoxazinone synthesis12* (*Bx12*), encoding an *O*-methyltransferase that catalyzes the formation of HDMBOA-Glc from DIMBOA-Glc (Meihls et al., 2013), is the most likely candidate gene for the common mQTL *1@85.5* (Figure 5E; Supplemental Figure 3; Supplemental Data Set 4). Consistent with this result, the content of HDMBOA and the expression level of *Bx12* in the maize-teosinte BC_2S_3 population were highly correlated ($r = 0.54$, $P < 2.2E-16$; Figures 5F and 5G). Meihls et al. (2013) identified a natural CACTA transposon insertion that inactivates *Bx12*. In our previous study, we further found that this CACTA transposon insertion was detected in only temperate maize and exhibited strong evidence of selection (Wang et al., 2018). Maize damage by aphids is a greater problem in temperate regions than in tropical regions (Ahmad et al., 2011; Meihls et al., 2013). DIMBOA-Glc content has been shown to be positively correlated with the level of resistance against maize leaf aphid (*Rhopalosiphum maidis*; Meihls et al., 2013; Betsiashvili et al., 2015; Zheng et al., 2015). Therefore, the CACTA transposon insertion that inactivates *Bx12* for accumulating more DIMBOA-Glc might be an important adaptive event for temperate maize to conquer the increased numbers of aphids in temperate environments. In addition to *Bx12*, several other benzoxazionoid genes, including *Bx7*, *Bx9*, and *Bx13*, colocalized with benzoxazinoid mQTLs (Figure 5B). These results suggested that these known *Bx* genes might play important roles in driving benzoxazinoid divergence.

Pr1 and *FHT1* Contributed to the Metabolic Divergence in the Flavonoid Pathway

Flavonoids constitute a large group of plant-specialized metabolites that play critical roles in plant development and defense (Andersen and Markham, 2006; Jiang et al., 2016). Flavonoids are synthesized from phenylpropanoid derivatives by condensation with malonyl-CoA (Figure 6A). These and further modifications of the heterocyclic ring yield a variety of structural forms, including flavones, flavanones, flavonols, isoflavones, 3-deoxy flavonoids, and anthocyanins (Andersen and Markham, 2006).

In this study, we identified 11 flavonoid compounds that exhibited altered contents in Teosinte-TroMaize or TroMaize-TemMaize (Figure 2E). In total, 73 mQTLs were mapped for the 11 flavonoid compounds (Figure 6B). Interestingly, mQTLs for different structural forms of flavonoid compounds colocalized in similar chromosomal positions (Figure 6B). For example, we detected a common mQTL for different forms of flavanones, flavones, isoflavones, and flavonols on chromosome 5 (*5@102.8*), which also corresponds to the mQTL hotspot *hs5-3* (Figures 4 and 6B). At this common mQTL (*5@102.8*), the 8759 allele is associated with higher contents of kaempferol, biochanin A derivatives, apigenin 4′-glucoside, puerarin 4′-*O*-glucoside, and chrysin derivative but lower levels of quercetin derivatives, tricetin derivatives, and coumarin derivative than the W22 allele (Figure 6B; Supplemental Data Set 3). Among the putative candidate genes identified for the common mQTL *5@102.8* (Supplemental Data Set 4), *Pr1*, a known gene encoding flavonoid 3′-hydroxylase (F3′H) that is responsible for catalyzing the conversion of naringenin to eriodictyol (Sharma et al., 2011), is the most likely candidate gene (Figure 6C). Interestingly, a significant *cis*-eQTL was detected for *Pr1*, with the peak SNP located 248 kb downstream of *Pr1* (Wang et al., 2018). At this *cis*-eQTL, the 8759 allele is associated with lower expression of *Pr1* than the W22 allele (Supplemental Figure 4). This *cis*-eQTL effect on *Pr1* expression is consistent with the mQTL effect on flavonoid contents, with high *Pr1* expression leading to increased levels of downstream metabolites of F3′H (Figures 6A and 6B). To further validate the role of *Pr1*, we obtained

Figure 6. (continued).

(B) mQTLs for the 11 divergent flavonoid compounds. Triangles represent the peak positions of the corresponding mQTL. Red triangles represent mQTLs at which the teosinte allele is associated with increased metabolite content relative to the maize allele. Blue triangles represent mQTLs at which the teosinte allele is associated with reduced metabolite content relative to the maize allele.

(C) Candidate gene analysis of the mQTL on chromosome 5 (*5@102.8*) for flavonoid. The original LOD profile of *5@102.8* is plotted in red. The LOD profiles of *5@102.8* from new models adding gene expression covariate are plotted in turquoise (significant covariates) and gray (nonsignificant covariates). The dark blue line represents *Pr1*.

(D) A *pr1* mutant containing a *Mu* transposon insertion in the first exon of the gene.

(E) Volcano plot of altered metabolites in the *pr1* mutant (only the 190 metabolites with annotated structure were plotted). Significantly altered metabolites were declared with fold change (FC) > 2 and $P < 0.01$ (Student's *t* test).

(F) Flavonoid profiles in seedlings of wild-type (WT) and *pr1* mutant plants. ***$P < 0.001$ (Student's *t* test, $n = 5$). Statistical analysis for the five replicates is shown in Supplemental Data Set 5.

(G) Candidate gene analysis for the mQTL on chromosome 2 (*2@12.9*) for tricetin derivative (m0157). The original LOD profile of *2@12.9* is plotted in red. The LOD profiles of *2@12.9* from new models adding gene expression covariate are plotted in turquoise (significant covariates) and gray (nonsignificant covariates). The dark blue line represents *FHT1*.

(H) Correlation between tricetin derivative (m0157) contents and *FHT1* expression levels in the maize-teosinte BC_2S_3 RIL population. RILs that are homozygous for 8759 and W22 and heterozygous at *FHT1* are indicated in red, blue, and green-yellow, respectively. The gene expression and metabolite content values are residuals after accounting for confounders.

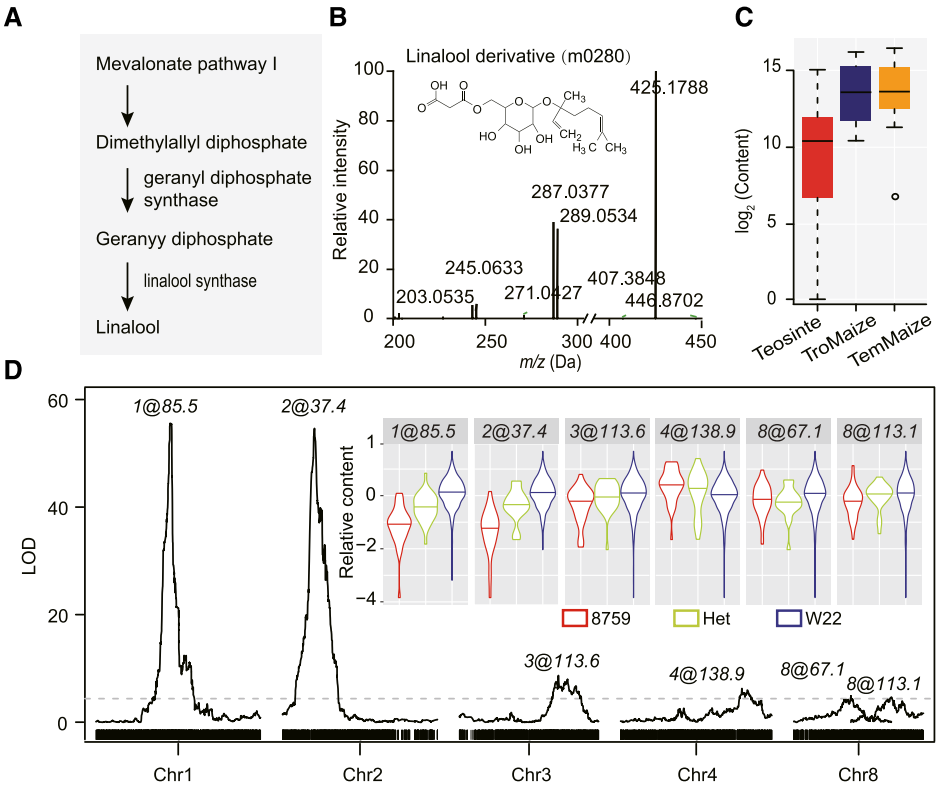

Figure 7. Genetic Basis of Linalool Derivative Divergence.

(A) Overview of linalool biosynthesis in maize. The enzyme geranyl diphosphate synthase catalyzes the synthesis of geranyl diphosphate from dimethylallyl diphosphate. Linalool synthase then catalyzes the conversion of geranyl diphosphate to linalool.
(B) Structure and LC-MS/MS fragmentation of the linalool 3-(6′′-malonylglucoside).
(C) Relative content of linalool 3-(6′′-malonylglucoside) in teosinte, tropical maize, and temperate maize. The *y* axis represents $\log_2$-transformed content.
(D) mQTL mapping for the linalool derivative in the maize-teosinte BC_2S_3 RIL population. The inset violin plots represent the allele effect for each mQTL. The red, blue, and green-yellow violin plots indicate RILs carrying homozygous 8759, homozygous W22, and heterozygous alleles at each mQTL in the maize-teosinte BC_2S_3 population, respectively. Chr, chromosome.

a transposon insertion mutant that contains a *Mu* transposon insertion in the first exon of *Pr1* (Figure 6D). We performed metabolic profiling for the *Mu* mutant and the wild-type using LC-HRMS. In total, 143 metabolites (7.8% of detected metabolites) were significantly altered ($P < 0.01$, fold change > 2) in the *pr1* mutant compared with the wild type (Figure 6E; Supplemental Data Set 5). Among these 143 altered metabolites, 26 had an annotated structure, including 20 flavonoids, 2 phenolic glycosides, 2 amino acids, 2 carboxylic acids, and 2 unclassified metabolites (Figure 6E). All of the eight metabolites that have mQTLs mapped at the *Pr1* locus exhibited significant changes in the *pr1* mutant (Figure 6F). These results further validated the role of *Pr1* in flavonoid biosynthesis.

Additionally, we found that eight mQTLs for different forms of flavanones, flavones, isoflavones, and flavonols that were targeted in Teosinte-TroMaize or TroMaize-TemMaize were located in similar positions on chromosome 2 (*2@12.9*), which also corresponds to the mQTL hotspot *hs2-1* (Figures 4 and 6B). At this common mQTL (*2@12.9*), the 8759 allele is associated with higher contents of kaempferol, quercetin derivative, and tricetin derivative but lower levels of apigenin derivative, chrysin derivative, biochanin A derivative, tricin derivative, and puerarin 4′-*O*-glucoside than the W22 allele (Figure 6B; Supplemental Data Set 3). Among the putative candidate genes (Supplemental Data Set 4), *FHT1* was identified as the most likely candidate for the common mQTL *2@12.9* (Figure 6G). *FHT1* encodes a flavanone 3-hydroxylase that is necessary for the production of flavonols and anthocyanins (Deboo et al., 1995). In our previous eQTL analysis (Wang et al., 2018), a significant *cis*-eQTL was detected for *FHT1*, with the peak SNP located 109 kb downstream of *FHT1*. At this *cis*-eQTL, the 8759 allele exhibited higher *FHT1* expression than the W22 allele (Figure 6H; Supplemental Figure 5). As expected, the *FHT1* expression level was significantly correlated with the tricetin derivative content in the maize-teosinte BC_2S_3 RIL population ($r = 0.36$, $P < 2.2E-16$; Figure 6H). Therefore, the *cis*-variant at *FHT1* might have played an important role in driving the metabolic divergence in flavonoid subgroups, including flavanones, flavones, isoflavones, and flavonols, since maize domestication.

ZmTPS1 Contributed to the Metabolic Divergence in the Terpenoid Pathway

Terpenoids form a large group of secondary metabolites that function in plant interactions with pollinators, herbivores, and

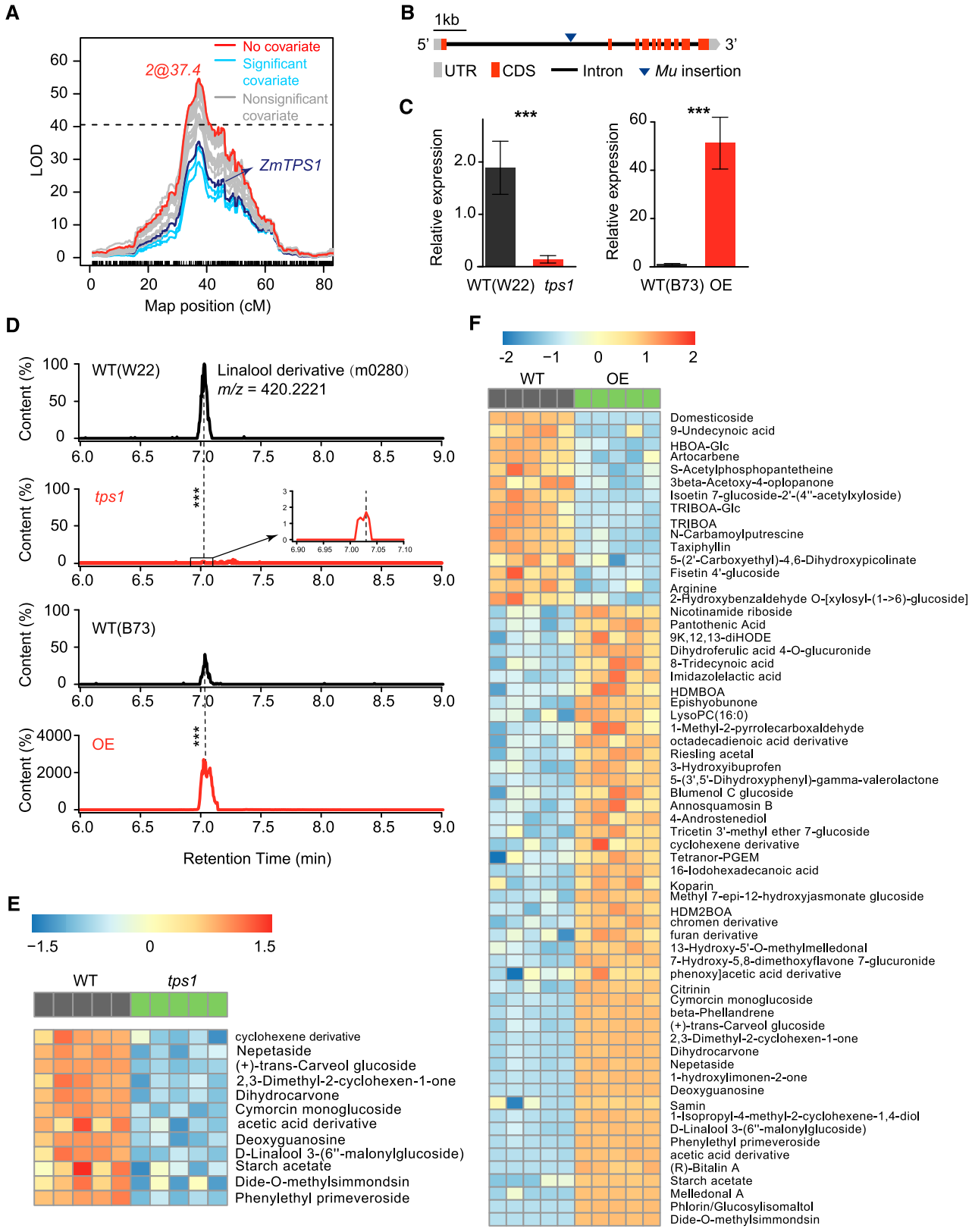

Figure 8. *ZmTPS1* Contributed to the Metabolic Divergence in the Terpenoid Pathway.

parasitoids (Kessler and Baldwin, 2001; Rector et al., 2003; Schnee et al., 2006; Richter et al., 2016). We found that five terpenoid compounds were specifically targeted in Teosinte-TroMaize and one was specifically targeted in TroMaize-TemMaize (Supplemental Figure 6). Linalool is a monoterpene that functions in various biological processes, including attraction of pollinators and resistance to biotic stresses (Raguso and Pichersky, 1999; van Schie et al., 2007). In our study, we identified a derivative of linalool [linalool 3-(6′′-malonylglucoside)], and its content was higher in tropical maize than in teosinte (Figures 7A–7C). Further linkage mapping in the maize-teosinte BC_2S_3 RIL population identified six mQTLs for the linalool derivative that jointly explained 54.1% of the phenotypic variation (Figure 7D; Supplemental Data Set 3). Relative to the 8759 allele, the W22 allele at five of the six loci conferred a higher level of the linalool derivative (Figure 7D), indicating their roles in increasing the content of the linalool derivative during the evolution from teosinte to tropical maize. Candidate gene analysis identified 59 putative candidate genes for the six linalool derivative mQTLs (Supplemental Data Set 4). Interestingly, *GRMZM2G049538* (*ZmTPS1*), a gene encoding a terpene synthase, is among the most likely candidates for the mQTL on chromosome 2 (*2@37.4*, Figure 8A). The amino acid sequence of *ZmTPS1* showed a high similarity to the terpene synthase sequences of the TPS-e/f subfamily (Supplemental Figure 7; Supplemental Data Set 6). This group includes *AtTPS04* in Arabidopsis (*Arabidopsis thaliana*), which is responsible for the formation of (*E*,*E*)-geranyllinalool from the substrate geranylgeranyl diphosphate (Herde et al., 2008). Indeed, a previous study has shown that ZmTPS1 could catalyze the formation of linalool in vitro (Schnee et al., 2002).

To further validate the function of *ZmTPS1*, we obtained a transposon insertion mutant that contains a *Mu* transposon insertion in the first intron of *ZmTPS1* (Figure 8B). At the same time, we overexpressed *ZmTPS1* under the control of the *Ubiquitin* promoter (Figure 8C). RT-qPCR analysis showed that *ZmTPS1* expression was reduced 13.5-fold in the *Mu* mutant but upregulated 51-fold in the overexpression line (Figure 8C). We performed metabolic profiling for the *Mu* mutant and the overexpression line using LC-HRMS. In the *tps1* mutant, 21 metabolites (1.2% of detected metabolites) exhibited altered contents compared with wild-type plants and 12 of them have annotated structure (Figures 8D and 8E; Supplemental Data Set 7). As expected, the level of linalool derivative was significantly reduced in the mutant (Figure 8D). In the *ZmTPS1* overexpression line, a total of 463 metabolites (25.5% of detected metabolites) exhibited altered contents compared with the wild type (Supplemental Data Set 7). Among these altered metabolites, 63 had annotated structures, mainly including 11 terpenoids, 12 carboxylic acids, 5 benzoxazinoids, and 5 flavonoids (Figure 8F; Supplemental Data Set 7). As expected, the linalool derivative content showed a dramatic increase in *ZmTPS1*-overexpressed plants (Figure 8D). Collectively, these results suggest that *ZmTPS1* is involved in terpenoid biosynthesis and might have a significant impact on other metabolite pathways.

Previous studies have shown that maize seedlings release large amounts of volatiles upon herbivore attack (Turlings et al., 1990, 1991). The volatile blend of maize plants is dominated by terpenes, including the monoterpene linalool and the sesquiterpenes (*E*)-β-farnesene and (*E*)-nerolidol (Turlings et al., 1991). These terpene volatiles often function as attractants for natural enemies of the herbivores, either predators or parasitoids, to locate their hosts or prey, consequently enhancing the defense of the host plants against herbivore attack (Turlings et al., 1990, 1991; Yan and Wang, 2006; Block et al., 2019). This defense mechanism, termed indirect defense, has been shown to play a critical role in a variety of plant species (Degenhardt, 2009; de Lange et al., 2014). Among the herbivore-induced volatile compounds, linalool was the most frequently occurring constitutive compound across different maize genotypes (Degen et al., 2004). Therefore, the increased linalool content in tropical maize compared with teosinte might be associated with an enhanced defense against herbivores in tropical maize. However, this hypothesis needs further experimental testing. Given the role of *ZmTPS1* in catalyzing the formation of linalool, we speculate that *ZmTPS1* may play an important role in regulating the divergence in linalool between teosinte and tropical maize.

METHODS

Plant Materials and Growth Conditions

The teosinte accessions included eight teosinte inbred lines provided by John Doebley (University of Wisconsin, Madison) and eight accessions from the Maize Coop Stock Center (Supplemental Table 1). The 27 maize inbred lines consisted of 12 tropical maize inbred lines and 15 temperate maize inbred lines that maximize the genetic diversity of maize (Supplemental Table 1; Liu et al., 2003; Yu et al., 2008). The 43 diverse maize and teosinte accessions were grown in a controlled greenhouse environment with 16 h of light (at 5000-lux illumination intensity and mixed wavelength ranging from 450 to 660 nm) and 8 h of dark. We used a randomized design with each genotype having at least two biological replicates.

Figure 8. (continued).

(A) Candidate gene analysis for the linalool derivative mQTL on chromosome 2 (*2@37.4*). The original LOD profile of *2@37.4* is plotted in red. The LOD profiles of *2@37.4* from new models adding gene expression covariate are plotted in turquoise (significant covariates) and gray (nonsignificant covariates). The dark blue line represents *ZmTPS1*.

(B) *tps1* mutant containing a *Mu* transposon insertion in the first intron of the gene. CDS, coding sequence; UTR, untranslated region.

(C) *ZmTPS1* expression levels in mutant (left) and overexpression (OE) plants (right). The expression values are relative to the control gene *ZmTubulin1* and represent means ± SD of five biological replicates. ***P < 0.001 (Student's *t* test). WT, wild type.

(D) Linalool 3-(6′′-malonylglucoside) profiles of wild-type (WT), *tps1* mutant, and overexpression (OE) plants. ***P < 0.001 (Student's *t* test, $n = 5$).

(E) Significantly altered metabolites in *tps1* mutant. The heatmap displays the 12 metabolites having annotated structure that exhibited significant difference between mutant and wild type (WT; fold change > 2 and P < 0.01).

(F) Significantly altered metabolites in *ZmTPS1* overexpressed plants. The heatmap displays the 63 metabolites with annotated structure that exhibited significant difference between overexpressed (OE) plants and wild-type (WT) plants (fold change > 2 and P < 0.01).

The BC_2S_3 RIL population, containing 624 lines, was derived from a cross between W22 (a typical temperate inbred line of *Zea mays* subsp *mays*) and CIMMYT accession 8759 (a typical accession of *Z. m.* subsp *parviglumis*). The population was obtained from the Maize Coop Stock Center and was previously genotyped using 19,838 SNPs (Shannon, 2012). The RIL population was grown in the same environment as the diverse maize and teosinte accessions with an augmented incomplete randomized block design. The planting of the BC_2S_3 RIL population has been described previously (Wang et al., 2018), and part of the sampled seedling tissues has been successfully used for transcriptome sequencing and eQTL analysis (Wang et al., 2018). The remaining samples were used for metabolite profiling in this study.

Sample Preparation

The aboveground seedlings were harvested 11 d after planting for metabolic profiling. The materials were snap frozen in liquid nitrogen, ground into powder, and partitioned into two sample sets for lipid-soluble and water-soluble metabolite extraction. One hundred milligrams of powder was weighed and transferred to a 1.5-mL centrifuge tube. One milliliter of absolute methanol was added, and the tube was placed in an ultrasonic cleaner (KunShan) for 30 min for lipid-soluble metabolite extraction (or 75% [v/v] methanol for water-soluble metabolites). After centrifugation (5417R centrifuge, Eppendorf) at 15,000 rpm and 4°C for 10 min, the supernatants of the extracts were combined (1:1, v/v), and 0.6 mL of the mixture was dried using a vacuum concentrator (Concentrator plus, Eppendorf). The dried samples were reconstituted in 80 μL of 50% (v/v) methanol and filtered through a 0.1-μm membrane for LC-HRMS analysis.

LC-HRMS Analysis

We performed ultrahigh performance LC-HRMS (UPLC-HRMS) analysis as described previously (Cao et al., 2016). Briefly, an UPLC-HRMS system (UPLC, ACQUITY i-Class, Waters; HRMS, Q-Exactive, Thermo Fisher Scientific) equipped with a heated electrospray ionization source was used for the analysis of metabolites in seedlings. MS analysis was performed in the positive ion mode, and we obtained scans in the mass range of *m/z* 70 to 1000 with a resolution of 70,000. To obtain sufficient data points for quantification, three scans were performed per second. The MS2 scan used a normalized collision energy of 35 V, an isolation window of 0.8 *m/z*, and a mass resolution of 35,000. Compounds were identified with accurate mass measurement of molecular ions and fragment ions with high resolution. We obtained a list of candidate chemical formulas after searching the accurate mass of molecular ions against compound databases such as Metabolite and Tandom MS Database (METLIN), Kyoto Encyclopedia of Genes and Genomes (KEGG), Plant Metabolic Pathway Databases (PlantCyc), Arabidopsis Metabolic Pathway Databases (AraCyc), Human Metabolome Database (HMDB), LIPID MAPS, High Resolution Mass Spectral Database (MassBank), PubChem, and MeSH with a mass accuracy of 5 ppm. The isotopic pattern of the molecular ions helped to determine the likely formulas. Tandem mass spectrometry (MS/MS) spectral database matches were used to match the fragment ion spectra to the candidate compounds. We also compared MS/MS spectra with theoretical fragmentation patterns with mass accuracy at 10 ppm using Progenesis QI 2.0 (Waters). Metabolite identifications were classified at level II based on their spectral similarities with public/commercial spectral libraries in accordance with the Metabolomics Standards Initiative guidelines (Sumner et al., 2007). Authentic standards for metabolites of particular interest were used to verify the identification results (Supplemental Data Set 1).

Statistical Analysis of Metabolite Variation

All the metabolite data were $\log_2$ transformed to improve normality for further analysis. For the diverse maize and teosinte accessions, least square means were calculated from a mixed linear model fitting genotype as a fixed effect and replicates and assaying batch as random effects. Broad sense heritability was estimated as:

$$H^2 = \sigma_g^2 \Big/ \left(\sigma_g^2 + \sigma_e^2 \Big/ n\right) \quad (1)$$

where σ_g^2 is the genetic variance, σ_e^2 is the residual error variance, and n is the number of replicates. Estimates of σ_g^2 and σ_e^2 were obtained from a mixed linear model with genotype, replicates, and assaying batch fitted as random effects. PCA of the metabolites across the 43 diverse maize and teosinte accessions was performed using the R statistical package (https://www.r-project.org/).

For the BC_2S_3 RIL population, we used a previously described strategy (Pickrell et al., 2010) to control the potential noise that could be introduced during the experiment. Following this strategy, each metabolite was fitted in a linear regression model to correct for known confounders (assaying batch, block in greenhouse) and unknown confounders (the first principal component from PCA). This correction approach has been widely applied in large-scale mapping studies and significantly increased statistical power (Pickrell et al., 2010; Wang et al., 2018). The metabolite contents after correction for non-genetic confounders were used for subsequent mQTL analysis.

Q_{ST}-F_{ST} Comparison

To identify divergent metabolites that resulted from selection rather than neutral processes, we used a Q_{ST}-F_{ST} comparison strategy (Leinonen et al., 2013). To estimate genetic differentiation, the 43 diverse maize and teosinte accessions were genotyped with a Maize6KSNP chip (Illumina). The data were filtered to remove SNPs that were monomorphic or had a missing rate greater than 10%. Finally, 3665 SNPs in total were used in the subsequent analysis.

We followed a previously described procedure to perform the Q_{ST}-F_{ST} comparison analysis (Beleggia et al., 2016). Briefly, F_{ST} was calculated using the method of Weir and Cockerham (1984) as follows:

$$F_{ST} = (H_T - H_S)/H_T \quad (2)$$

where H_T refers to heterozygosity in overall population and H_T refers to heterozygosity within subpopulations. To estimate the level of neutral population divergence, we first excluded SNPs that deviate from neutral expectations using an F_{ST}-based outlier test (Beleggia et al., 2016). The outliers were identified by permuting individuals across populations. The outlier SNPs were discarded, and the final set of neutral SNPs was used to calculate neutral F_{ST}. Finally, 2874 and 3236 neutral SNPs were used to calculate the neutral F_{ST} of Teosinte-TroMaize and TroMaize-TemMaize, respectively.

We calculated Q_{ST} for each metabolite as follows (Bonnin et al., 1996; Beleggia et al., 2016):

$$Q_{ST} = (1 + F_{IS})\sigma_B^2 \Big/ \left[(1 + F_{IS})\sigma_B^2 + 2\sigma_W^2\right] \quad (3)$$

where σ_B^2 is the between-population variance component, σ_W^2 is the within-population variance component, and F_{IS} is the inbreeding coefficient. GCTA (Yang et al., 2011) was used to compute the inbreeding coefficient. We estimated 99% confidence intervals (CIs) of Q_{ST} by resampling individuals with replacement 1000 times (Evans et al., 2014). We compared the 99% CIs of Q_{ST} with neutral F_{ST} values, and those metabolites with CIs greater than the neutral F_{ST} value were considered targets of selection.

mQTL Mapping

Following the previously described procedure (Shannon, 2012; Huang et al., 2016, 2018; Li et al., 2016; Xu et al., 2017a, 2017b; Wang et al., 2018), we performed mQTL mapping using a modified version of R/qtl (Broman et al., 2003) that considers the BC_2S_3 pedigree of the RILs (Shannon, 2012).

The analysis code is available in GitHub (https://github.com/Guanghui-Xu/mQTL_code). Briefly, mQTL mapping was first performed using the Haley–Knott regression method to generate an initial list of mQTLs (Broman et al., 2003). Permutations (1000 times) were conducted to determine the significance threshold of LOD for claiming significant mQTLs. A multiple mQTL model was then fitted to test the significance of each mQTL using a drop-one analysis of variance (ANOVA). The positions of mQTLs remained in the multiple mQTL model were further refined using the *refineqtl* command. The total phenotypic variation explained by all the mQTLs detected for a metabolite was determined by fitting all mQTL terms in the model. The phenotypic variation explained by each mQTL was calculated using the drop-one ANOVA with the *fitqtl* function. The confidence interval for each mQTL was defined using a 2-LOD drop support interval.

To identify potential mQTL hotspots, we performed a sliding window analysis with 3-centimorgan (cM) windows and 1-cM steps to count the number of mQTL in each window according to the peak positions of the mQTLs. Following a previously described procedure (Wang et al., 2018), we performed permutation tests (1000 times) to determine the threshold for declaring significant mQTL hotspots. In each permutation test, the 1494 mQTLs were randomly assigned to windows in the genome, and the number of mQTLs in each window was counted. The maximal number of mQTLs across the windows in each permutation was recorded. On the basis of the distribution of the maximal mQTL number from permutations, the threshold for declaring an mQTL hotspot was seven mQTLs per window at a significant threshold of $P = 0.05$. Adjacent windows that exceeded the hotspot threshold were merged. The pathway enrichment analysis was performed using a hypergeometric test, and the false discovery rate method was used for multiple test correction.

Candidate Gene Analysis

Using the same seedling tissues, we previously performed transcriptome sequencing and conducted eQTL analysis in the maize-teosinte BC_2S_3 RIL population (Wang et al., 2018). This data set enabled us to integrate the transcriptome to identify specific genes whose expression regulates metabolite variation in the population. We used a previously described candidate gene ranking approach to identify the candidate genes for mQTLs (Lovell et al., 2015). The analysis code is available in GitHub (https://github.com/Guanghui-Xu/mQTL_code). For each mQTL, the expression of each expressed gene within its 2-LOD support interval was added as a covariate to the original multiple-mQTL model to test the impact of gene expression on the identification of the target mQTL (Supplemental Figure 2). The difference in mQTL LOD scores (designated $\triangle$LOD) between the original multiple-mQTL model and the new model with a gene-expression covariate was used as a statistic to measure the relative effect of the candidate gene on the target mQTL (Supplemental Figure 2). To determine the significance of $\triangle$LOD, a permutation test was performed for each mQTL. For each mQTL, the expression of a randomly selected gene was added as covariate to the original multiple mQTL model to estimate the impact of gene expression on target mQTL by chance alone. We repeated this procedure 10,000 times to generate a simulated ΔLOD distribution, based on which we determined the significance of ΔLOD at $P = 0.001$.

Gene Expression Quantification by RT-qPCR

To examine the expression level of *ZmTPS1* in the *Mu* transposon insertion mutant and the overexpression line, the *tps1* mutant, the T3 overexpression plants, and their corresponding wild-type plants were grown in a greenhouse. Seedling tissues were harvested 11 d after planting, with five biological replicates for each genotype. Total RNA was extracted using the RNAprep Pure Plant Kit (Tiangen) according to the manufacturer's instructions. The RNA was treated with RNase-free DNase I (Takara), and its concentration and quality were assessed using a NanoDrop ND-1000 spectrophotometer (Thermo Fisher Scientific). Next, 2 μg of purified RNA was reverse transcribed using Moloney murine leukemia virus reverse transcriptase (Promega) and random primers. RT-qPCR was performed on an ABI 7500 instrument (Applied Biosystems) using the SYBR Premix Ex Taq II kit (Takara), and measurements were obtained using the relative quantification method (Schmittgen and Livak, 2008). The tubulin gene *ZmTubulin1* (GRMZM2G152466) was used as a reference gene. Primers for the reference gene and *ZmTPS1* are shown in Supplemental Table 3.

Phylogenetic Analysis

Protein sequences of terpene synthases from maize and other plant species were obtained from Gramene (http://www.gramene.org), The Institute for Genomic Research (http://rice.plantbiology.msu.edu/index.shtml), and The Arabidopsis Information Resource (http://www.arabidopsis.org) databases. Protein sequences were aligned with MUSCLE (Edgar, 2004) and manually edited when necessary. A neighbor-joining tree was constructed using MEGA6.0 software with 1000 bootstrap replicates (Tamura et al., 2013).

Transgenic Functional Validation

The coding region of *ZmTPS1* was amplified from B73 cDNA and cloned into the binary vector pCUNm under the control of the *Ubiquitin* promoter. This construct was introduced into the receptor line B73 via *Agrobacterium tumefaciens*–mediated transformation (Ishida et al., 2007). The transgenic seeds were created by the maize functional genomics project of China Agricultural University. Transgene-positive and transgene-negative plants in *ZmTPS1* transgenic families were identified using transgene-specific primers (Supplemental Table 3). Seedlings of T3 transgene-positive and transgene-negative plants were sampled for gene expression and metabolite profiling. Five biological replicates were used for each genotype. Quantification and annotation of compounds were performed as described previously. Metabolites that exhibited significant alteration in overexpression plants were identified using Student's *t* test, with a significance threshold of $P < 0.01$ and fold change > 2.

Mutant Functional Validation

Plants with *Mu* insertions in *Pr1* (UFMu-00761) and *ZmTPS1* (UFMu-06794) were obtained from the Maize Coop Stock Center and confirmed using a combination of gene- and transposon-specific primers (Supplemental Table 3). *Mu*-containing plants were used as donor parents and backcrossed to the wild-type W22 two times, followed by self-pollination to generate a segregating population. The homozygous mutant and wild-type plants in the segregating population were used for expression and metabolite quantification. Five biological replicates were used for each genotype. The significantly altered metabolites in the mutants were identified using Student's *t* test, with a significance threshold of $P < 0.01$ and fold change > 2 (Supplemental Data Sets 5 and 7).

Accession Numbers

Sequence data from this article can be found in the GenBank/EMBL data libraries under accession numbers *Bx12*, GRMZM2G023325; *Pr1*, GRMZM2G025832; *FHT1*, GRMZM2G062396; *ZmTPS1*, GRMZM2G049538.

Supplemental Data

Supplemental Figure 1. Principal component analysis (PCA) of the maize and teosinte accessions with 291 annotated metabolites.

Supplemental Figure 2. Pipeline for candidate gene ranking analysis.

Supplemental Figure 3. *Bx12* contributes to DIMBOA natural variation in the maize-teosinte BC_2S_3 RIL population.

Supplemental Figure 4. *cis*-eQTL for *Pr1* detected in the maize-teosinte BC_2S_3 RIL population.

Supplemental Figure 5. *cis*-eQTL for *FHT1* detected in the maize-teosinte BC_2S_3 RIL population.

Supplemental Figure 6. Divergent metabolites in the terpenoid pathway.

Supplemental Figure 7. Phylogenetic tree of the terpene synthase genes homologous to *ZmTPS1*.

Supplemental Table 1. Maize and teosinte materials used in this study.

Supplemental Table 2. mQTL hotspots and pathway enrichment analysis.

Supplemental Table 3. Primers used in this study.

Supplemental Data Set 1. Detailed information of the 291 annotated metabolites detected in this study.

Supplemental Data Set 2. Metabolite content in the maize-teosinte BC_2S_3 RIL population.

Supplemental Data Set 3. mQTLs for the divergent metabolites.

Supplemental Data Set 4. Candidate genes for mQTLs identified through candidate gene ranking analysis.

Supplemental Data Set 5. Detailed information for the altered metabolites in the *pr1* mutant.

Supplemental Data Set 6. Amino acid sequence alignments of the terpene synthase genes homologous to *ZmTPS1*.

Supplemental Data Set 7. Detailed information for the altered metabolites in *tps1* mutant and overexpression line.

ACKNOWLEDGMENTS

We thank John Doebley for critically reading the manuscript. This research was supported by the National Key Research and Development Program of China (2016YFD0100303), the National Natural Science Foundation of China (31421005), the Recruitment Program of Global Experts, and the Fundamental Research Funds for the Central Universities.

AUTHOR CONTRIBUTIONS

G.X. and F.T. designed the research. J.C. and G.X. performed metabolic profiling. G.X. collected and analyzed the data. G.X., X.W., Q.C., W.J., Z.L., and F.T. wrote and edited the article. All authors read and approved the final article.

Received February 19, 2019; revised June 3, 2019; accepted June 17, 2019; published June 21, 2019.

REFERENCES

Ahmad, S., Veyrat, N., Gordon-Weeks, R., Zhang, Y., Martin, J., Smart, L., Glauser, G., Erb, M., Flors, V., Frey, M., and Ton, J. (2011). Benzoxazinoid metabolites regulate innate immunity against aphids and fungi in maize. Plant Physiol. **157:** 317–327.

Alseekh, S., Ofner, I., Pleban, T., Tripodi, P., Di Dato, F., Cammareri, M., Mohammad, A., Grandillo, S., Fernie, A.R., and Zamir, D. (2013). Resolution by recombination: Breaking up *Solanum pennellii* introgressions. Trends Plant Sci. **18:** 536–538.

Andersen, O.M., and Markham, K.R. (2006). Flavonoids: Chemistry, Biochemistry and Applications. (Boca Raton, FL: CRC Press).

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.

Beleggia, R., Rau, D., Laidò, G., Platani, C., Nigro, F., Fragasso, M., De Vita, P., Scossa, F., Fernie, A.R., Nikoloski, Z., and Papa, R. (2016). Evolutionary metabolomics reveals domestication-associated changes in tetraploid wheat kernels. Mol. Biol. Evol. **33:** 1740–1753.

Betsiashvili, M., Ahern, K.R., and Jander, G. (2015). Additive effects of two quantitative trait loci that confer *Rhopalosiphum maidis* (corn leaf aphid) resistance in maize inbred line Mo17. J. Exp. Bot. **66:** 571–578.

Block, A.K., Vaughan, M.M., Schmelz, E.A., and Christensen, S.A. (2019). Biosynthesis and function of terpenoid defense compounds in maize (*Zea mays*). Planta **249:** 21–30.

Bonnin, I., Prosperi, J.M., and Olivieri, I. (1996). Genetic markers and quantitative genetic variation in *Medicago truncatula* (Leguminosae): A comparative analysis of population structure. Genetics **143:** 1795–1805.

Broman, K.W., Wu, H., Sen, S., and Churchill, G.A. (2003). R/qtl: QTL mapping in experimental crosses. Bioinformatics **19:** 889–890.

Cao, J., Li, M., Chen, J., Liu, P., and Li, Z. (2016). Effects of MeJA on Arabidopsis metabolome under endogenous JA deficiency. Sci. Rep. **6:** 37674.

Chan, E.K.F., 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.

Chen, W., et al. (2014). Genome-wide association analyses provide genetic and biochemical insights into natural variation in rice metabolism. Nat. Genet. **46:** 714–721.

Chen, W., et al. (2016). Comparative and parallel genome-wide association studies for metabolic and agronomic traits in cereals. Nat. Commun. **7:** 12767.

Chen, Y., Lee, L.S., Luckett, D.J., Henry, R., Hill, H., and Edwards, M. (2007). A quinolizidine alkaloid O-tigloyltransferase gene in wild and domesticated white lupin (*Lupinus albus*). Ann. Appl. Biol. **151:** 357–362.

Deboo, G.B., Albertsen, M.C., and Taylor, L.P. (1995). Flavanone 3-hydroxylase transcripts and flavonol accumulation are temporally coordinate in maize anthers. Plant J. **7:** 703–713.

Degen, T., Dillmann, C., Marion-Poll, F., and Turlings, T.C.J. (2004). High genetic variability of herbivore-induced volatile emission within a broad range of maize inbred lines. Plant Physiol. **135:** 1928–1938.

Degenhardt, J. (2009). Indirect defense responses to herbivory in grasses. Plant Physiol. **149:** 96–102.

de Lange, E.S., 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.

Dixon, R.A., and Strack, D. (2003). Phytochemistry meets genome analysis, and beyond. Phytochemistry **62:** 815–816.

Doebley, J. (2004). The genetics of maize evolution. Annu. Rev. Genet. **38:** 37–59.

Edgar, R.C. (2004). MUSCLE: Multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. **32:** 1792–1797.

Enneking, D., and Wink, M. (2000). Towards the elimination of anti-nutritional factors in grain legumes. In Linking Research and Marketing Opportunities for Pulses in the 21st Century, R. Knight, ed (Dordrecht, The Netherlands: Kluwer Academic Publishers), pp. 671–683.

Evans, L.M., Slavov, G.T., Rodgers-Melnick, E., Martin, J., Ranjan, P., Muchero, W., Brunner, A.M., Schackwitz, W., Gunter, L., Chen, J.-G., Tuskan, G.A., and DiFazio, S.P. (2014). Population genomics of *Populus trichocarpa* identifies signatures of selection and adaptive trait associations. Nat. Genet. **46:** 1089–1096.

Fang, C., Fernie, A.R., and Luo, J. (2019). Exploring the diversity of plant metabolism. Trends Plant Sci. **24:** 83–98.

Fernie, A.R., and Tohge, T. (2017). The genetics of plant metabolism. Annu. Rev. Genet. **51:** 287–310.

Frick, K.M., Kamphuis, L.G., Siddique, K.H.M., Singh, K.B., and Foley, R.C. (2017). Quinolizidine alkaloid biosynthesis in lupins and prospects for grain quality improvement. Front. Plant Sci. **8:** 87.

Glauser, G., Marti, G., Villard, N., Doyen, G.A., Wolfender, J.-L., Turlings, T.C.J., and Erb, M. (2011). Induction and detoxification of maize 1,4-benzoxazin-3-ones by insect herbivores. Plant J. **68:** 901–911.

Gong, L., Chen, W., Gao, Y., Liu, X., Zhang, H., Xu, C., Yu, S., Zhang, Q., and Luo, J. (2013). Genetic analysis of the metabolome exemplified using a rice population. Proc. Natl. Acad. Sci. USA **110:** 20320–20325.

Guo, L., et al. (2018). Stepwise *cis*-regulatory changes in *ZCN8* contribute to maize flowering-time adaptation. Curr. Biol. **28:** 3005–3015.e4.

Handrick, V., et al. (2016). Biosynthesis of 8-O-methylated benzoxazinoid defense compounds in maize. Plant Cell **28:** 1682–1700.

Herde, M., Gärtner, K., Köllner, T.G., Fode, B., Boland, W., Gershenzon, J., Gatz, C., and Tholl, D. (2008). Identification and regulation of TPS04/GES, an Arabidopsis geranyllinalool synthase catalyzing the first step in the formation of the insect-induced volatile C16-homoterpene TMTT. Plant Cell **20:** 1152–1168.

Hill, C.B., Taylor, J.D., Edwards, J., Mather, D., Langridge, P., Bacic, A., and Roessner, U. (2015). Detection of QTL for metabolic and agronomic traits in wheat with adjustments for variation at genetic loci that affect plant phenology. Plant Sci. **233:** 143–154.

Huang, C., et al. (2018). *ZmCCT9* enhances maize adaptation to higher latitudes. Proc. Natl. Acad. Sci. USA **115:** E334–E341.

Huang, C., Chen, Q., Xu, G., Xu, D., Tian, J., and Tian, F. (2016). Identification and fine mapping of quantitative trait loci for the number of vascular bundle in maize stem. J. Integr. Plant Biol. **58:** 81–90.

Hufford, M.B., et al. (2012). Comparative population genomics of maize domestication and improvement. Nat. Genet. **44:** 808–811.

Hung, H.-Y., Shannon, L.M., Tian, F., Bradbury, P.J., Chen, C., Flint-Garcia, S.A., McMullen, M.D., Ware, D., Buckler, E.S., Doebley, J.F., and Holland, J.B. (2012). *ZmCCT* and the genetic basis of day-length adaptation underlying the postdomestication spread of maize. Proc. Natl. Acad. Sci. USA **109:** E1913–E1921.

Ishida, Y., Hiei, Y., and Komari, T. (2007). *Agrobacterium*-mediated transformation of maize. Nat. Protoc. **2:** 1614–1621.

Jiang, N., Doseff, A.I., and Grotewold, E. (2016). Flavones: From biosynthesis to health benefits. Plants (Basel) **5:** 27.

Johns, T., and Alonso, J.G. (1990). Glycoalkaloid change during the domestication of the potato, *Solanum* Section *Petota.* Euphytica **50:** 203–210.

Kessler, A., and Baldwin, I.T. (2001). Defensive function of herbivore-induced plant volatile emissions in nature. Science **291:** 2141–2144.

Keurentjes, J.J.B. (2009). Genetical metabolomics: Closing in on phenotypes. Curr. Opin. Plant Biol. **12:** 223–230.

Keurentjes, J.J.B., Fu, J., de Vos, C.H.R., Lommen, A., Hall, R.D., Bino, R.J., van der Plas, L.H.W., Jansen, R.C., Vreugdenhil, D., and Koornneef, M. (2006). The genetics of plant metabolism. Nat. Genet. **38:** 842–849.

Knoch, D., Riewe, D., Meyer, R.C., Boudichevskaia, A., Schmidt, R., and Altmann, T. (2017). Genetic dissection of metabolite variation in Arabidopsis seeds: Evidence for mQTL hotspots and a master regulatory locus of seed metabolism. J. Exp. Bot. **68:** 1655–1667.

Leinonen, T., McCairns, R.J.S., O'Hara, R.B., and Merilä, J. (2013). Q(ST)-F(ST) comparisons: Evolutionary and ecological insights from genomic heterogeneity. Nat. Rev. Genet. **14:** 179–190.

Lemmon, Z.H., Bukowski, R., Sun, Q., and Doebley, J.F. (2014). The role of *cis* regulatory evolution in maize domestication. PLoS Genet. **10:** e1004745.

Li, D., et al. (2016). The genetic architecture of leaf number and its genetic relationship to flowering time in maize. New Phytol. **210:** 256–268.

Li, H., et al. (2013). Genome-wide association study dissects the genetic architecture of oil biosynthesis in maize kernels. Nat. Genet. **45:** 43–50.

Liang, Y., et al. (2019). *ZmMADS69* functions as a flowering activator through the *ZmRap2.7-ZCN8* regulatory module and contributes to maize flowering time adaptation. New Phytol. **221:** 2335–2347.

Liu, H., Wang, X., Warburton, M.L., Wen, W., Jin, M., Deng, M., Liu, J., Tong, H., Pan, Q., Yang, X., and Yan, J. (2015). Genomic, transcriptomic, and phenomic variation reveals the complex adaptation of modern maize breeding. Mol. Plant **8:** 871–884.

Liu, K., Goodman, M., Muse, S., Smith, J.S., Buckler, E., and Doebley, J. (2003). Genetic structure and diversity among maize inbred lines as inferred from DNA microsatellites. Genetics **165:** 2117–2128.

Lovell, J.T., Mullen, J.L., Lowry, D.B., Awole, K., Richards, J.H., Sen, S., Verslues, P.E., Juenger, T.E., and McKay, J.K. (2015). Exploiting differential gene expression and epistasis to discover candidate genes for drought-associated QTLs in *Arabidopsis thaliana.* Plant Cell **27:** 969–983.

Maresh, J., Zhang, J., and Lynn, D.G. (2006). The innate immunity of maize and the dynamic chemical strategies regulating two-component signal transduction in *Agrobacterium tumefaciens.* ACS Chem. Biol. **1:** 165–175.

Matsuda, F., Okazaki, Y., Oikawa, A., Kusano, M., Nakabayashi, R., Kikuchi, J., Yonemaru, J., Ebana, K., Yano, M., and Saito, K. (2012). Dissection of genotype-phenotype associations in rice grains using metabolome quantitative trait loci analysis. Plant J. **70:** 624–636.

Matsuda, F., Nakabayashi, R., Yang, Z., Okazaki, Y., Yonemaru, J., Ebana, K., Yano, M., and Saito, K. (2015). Metabolome-genome-wide association study dissects genetic architecture for generating natural variation in rice secondary metabolism. Plant J. **81:** 13–23.

Matsuoka, Y., Vigouroux, Y., Goodman, M.M., Sanchez G, J., Buckler, E., and Doebley, J. (2002). A single domestication for maize shown by multilocus microsatellite genotyping. Proc. Natl. Acad. Sci. USA **99:** 6080–6084.

Meihls, L.N., Handrick, V., Glauser, G., Barbier, H., Kaur, H., Haribal, M.M., Lipka, A.E., Gershenzon, J., Buckler, E.S., Erb, M., Köllner, T.G., and Jander, G. (2013). Natural variation in maize aphid resistance is associated with 2,4-dihydroxy-7-methoxy-1,4-benzoxazin-3-one glucoside methyltransferase activity. Plant Cell **25:** 2341–2355.

Niemeyer, H.M. (1988). Hydroxamic acids (4-hydroxy-1,4-benzoxazin-3-ones), defence chemicals in the Gramineae. Phytochemistry **27:** 3349–3358.

Pickrell, J.K., Marioni, J.C., Pai, A.A., Degner, J.F., Engelhardt, B.E., Nkadori, E., Veyrieras, J.-B., Stephens, M., Gilad, Y., and Pritchard, J.K. (2010). Understanding mechanisms underlying human gene expression variation with RNA sequencing. Nature **464:** 768–772.

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.

Raguso, R.A., and Pichersky, E. (1999). A day in the life of a linalool molecule: Chemical communication in a plant-pollinator system. Part 1: Linalool biosynthesis in flowering plants. Plant Species Biol. **14:** 95–120.

Rambla, J.L., et al. (2017). Identification, introgression, and validation of fruit volatile QTLs from a red-fruited wild tomato species. J. Exp. Bot. **68:** 429–442.

Rector, B.G., Liang, G., and Guo, Y. (2003). Effect of maysin on wild-type, deltamethrin-resistant, and Bt-resistant *Helicoverpa armigera* (Lepidoptera: Noctuidae). J. Econ. Entomol. **96:** 909–913.

Richter, A., et al. (2016). Characterization of biosynthetic pathways for the production of the volatile homoterpenes DMNT and TMTT in *Zea mays.* Plant Cell **28:** 2651–2665.

Riedelsheimer, 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. USA **109:** 8872–8877.

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.

Schauer, N., et al. (2006). Comprehensive metabolic profiling and phenotyping of interspecific introgression lines for tomato improvement. Nat. Biotechnol. **24:** 447–454.

Schmittgen, T.D., and Livak, K.J. (2008). Analyzing real-time PCR data by the comparative C(T) method. Nat. Protoc. **3:** 1101–1108.

Schnee, C., Köllner, T.G., Gershenzon, J., and Degenhardt, J. (2002). The maize gene *terpene synthase 1* encodes a sesquiterpene synthase catalyzing the formation of (E)-beta-farnesene, (E)-nerolidol, and (E,E)-farnesol after herbivore damage. Plant Physiol. **130:** 2049–2060.

Schnee, C., Köllner, T.G., Held, M., Turlings, T.C.J., Gershenzon, J., and Degenhardt, J. (2006). The products of a single maize sesquiterpene synthase form a volatile defense signal that attracts natural enemies of maize herbivores. Proc. Natl. Acad. Sci. USA **103:** 1129–1134.

Schwab, W. (2003). Metabolome diversity: Too few genes, too many metabolites? Phytochemistry **62:** 837–849.

Shang, Y., et al. (2014). Plant science. Biosynthesis, regulation, and domestication of bitterness in cucumber. Science **346:** 1084–1088.

Shannon, L. (2012). The Genetic Architecture of Maize Domestication and Range Expansion. PhD dissertation (Madison: University of Wisconsin).

Sharma, M., Cortes-Cruz, M., Ahern, K.R., McMullen, M., Brutnell, T.P., and Chopra, S. (2011). Identification of the *pr1* gene product completes the anthocyanin biosynthesis pathway of maize. Genetics **188:** 69–79.

Studer, A., Zhao, Q., Ross-Ibarra, J., and Doebley, J. (2011). Identification of a functional transposon insertion in the maize domestication gene *tb1.* Nat. Genet. **43:** 1160–1163.

Sumner, L.W., et al. (2007). Proposed minimum reporting standards for chemical analysis Chemical Analysis Working Group (CAWG) Metabolomics Standards Initiative (MSI). Metabolomics **3:** 211–221.

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.

Tamura, K., Stecher, G., Peterson, D., Filipski, A., and Kumar, S. (2013). MEGA6: Molecular evolutionary genetics analysis version 6.0. Mol. Biol. Evol. **30:** 2725–2729.

Tieman, D., et al. (2017). A chemical genetic roadmap to improved tomato flavor. Science **355:** 391–394.

Toubiana, D., Semel, Y., Tohge, T., Beleggia, R., Cattivelli, L., Rosental, L., Nikoloski, Z., Zamir, D., Fernie, A.R., and Fait, A. (2012). Metabolic profiling of a mapping population exposes new insights in the regulation of seed metabolism and seed, fruit, and plant relations. PLoS Genet. **8:** e1002612.

Turlings, T.C.J., Tumlinson, J.H., and Lewis, W.J. (1990). Exploitation of herbivore-induced plant odors by host-seeking parasitic wasps. Science **250:** 1251–1253.

Turlings, T.C.J., Tumlinson, J.H., Heath, R.R., Proveaux, A.T., and Doolittle, R.E. (1991). Isolation and identification of allelochemicals that attract the larval parasitoid, *Cotesia marginiventris* (Cresson), to the microhabitat of one of its hosts. J. Chem. Ecol. **17:** 2235–2251.

Tzin, V., Lindsay, P.L., Christensen, S.A., Meihls, L.N., Blue, L.B., and Jander, G. (2015). Genetic mapping shows intraspecific variation and transgressive segregation for caterpillar-induced aphid resistance in maize. Mol. Ecol. **24:** 5739–5750.

van Schie, C.C.N., Haring, M.A., and Schuurink, R.C. (2007). Tomato linalool synthase is induced in trichomes by jasmonic acid. Plant Mol. Biol. **64:** 251–263.

Verpoorte, R., and Memelink, J. (2002). Engineering secondary metabolite production in plants. Curr. Opin. Biotechnol. **13:** 181–187.

Wallace, J.G., Larsson, S.J., and Buckler, E.S. (2014). Entering the second century of maize quantitative genetics. Heredity **112:** 30–38.

Wang, H., Nussbaum-Wagler, T., Li, B., Zhao, Q., Vigouroux, Y., Faller, M., Bomblies, K., Lukens, L., and Doebley, J.F. (2005). The origin of the naked grains of maize. Nature **436:** 714–719.

Wang, X., Chen, Q., Wu, Y., Lemmon, Z.H., Xu, G., Huang, C., Liang, Y., Xu, D., Li, D., Doebley, J.F., and Tian, F. (2018). Genome-wide analysis of transcriptional variability in a large maize-teosinte population. Mol. Plant **11:** 443–459.

Weir, B.S., and Cockerham, C.C. (1984). Estimating *F*-statistics for the analysis of population structure. Evolution **38:** 1358–1370.

Wen, W., 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.

Wen, W., Li, D., Li, X., Gao, Y., Li, W., Li, H., Liu, J., Liu, H., Chen, W., Luo, J., and Yan, J. (2014). Metabolome-based genome-wide association study of maize kernel leads to novel biochemical insights. Nat. Commun. **5:** 3438.

Westhues, M., et al. (2017). Omics-based hybrid prediction in maize. Theor. Appl. Genet. **130:** 1927–1939.

Wills, D.M., Whipple, C.J., Takuno, S., Kursel, L.E., Shannon, L.M., Ross-Ibarra, J., and Doebley, J.F. (2013). From many, one: Genetic control of prolificacy during maize domestication. PLoS Genet. **9:** e1003604.

Wu, S., Tohge, T., Cuadros-Inostroza, Á., Tong, H., Tenenboim, H., Kooke, R., Méret, M., Keurentjes, J.B., Nikoloski, Z., Fernie, A.R., Willmitzer, L., and Brotman, Y. (2018). Mapping the Arabidopsis metabolic landscape by untargeted metabolomics at different environmental conditions. Mol. Plant **11:** 118–134.

Xu, D., et al. (2017a). *Glossy15* plays an important role in the divergence of the vegetative transition between maize and its progenitor, teosinte. Mol. Plant **10:** 1579–1583.

Xu, G., Wang, X., Huang, C., Xu, D., Li, D., Tian, J., Chen, Q., Wang, C., Liang, Y., Wu, Y., Yang, X., and Tian, F. (2017b). Complex genetic architecture underlies maize tassel domestication. New Phytol. **214:** 852–864.

Yan, Z.G., and Wang, C.Z. (2006). Identification of *Mythmna separata*-induced maize volatile synomones that attract the parasitoid *Campoletis chlorideae.* J. Appl. Entomol. **130:** 213–219.

Yang, Q., et al. (2013). CACTA-like transposable element in *ZmCCT* attenuated photoperiod sensitivity and accelerated the postdomestication spread of maize. Proc. Natl. Acad. Sci. USA **110:** 16969–16974.

Yang, J., Lee, S.H., Goddard, M.E., and Visscher, P.M. (2011). GCTA: A tool for genome-wide complex trait analysis. Am. J. Hum. Genet. **88:** 76–82.

Yu, J., Holland, J.B., McMullen, M.D., and Buckler, E.S. (2008). Genetic design and statistical power of nested association mapping in maize. Genetics **178:** 539–551.

Zheng, L., McMullen, M.D., Bauer, E., Schön, C.C., Gierl, A., and Frey, M. (2015). Prolonged expression of the BX1 signature enzyme is associated with a recombination hotspot in the benzoxazinoid gene cluster in *Zea mays.* J. Exp. Bot. **66:** 3917–3930.

Zhou, Y., et al. (2016). Convergence and divergence of bitterness biosynthesis and regulation in Cucurbitaceae. Nat. Plants **2:** 16183.

Zhu, G., et al. (2018). Rewiring of the fruit metabolome in tomato breeding. Cell **172:** 249–261.e12.