Genomic, Transcriptomic, and Phenomic Variation Reveals the Complex Adaptation of Modern Maize Breeding

Haijun Liu[1,4], Xiaqing Wang[1,4], Marilyn L. Warburton[2], Weiwei Wen[1], Minliang Jin[1], Min Deng[1], Jie Liu[1], Hao Tong[1], Qingchun Pan[1], Xiaohong Yang[3] and Jianbing Yan[1,*]

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

[2]Corn Host Plant Resistance Research Unit, United States Department of Agriculture-Agricultural Research Service, Box 9555, Mississippi State, MS 39762, USA

[3]National Maize Improvement Center of China, Beijing Key Laboratory of Crop Genetic Improvement, China Agricultural University, Beijing 100193, China

[4]These authors contributed equally to this article.

*Correspondence: Jianbing Yan (yjianbing@mail.hzau.edu.cn)

http://dx.doi.org/10.1016/j.molp.2015.01.016

ABSTRACT

The temperate-tropical division of early maize germplasms to different agricultural environments was arguably the greatest adaptation process associated with the success and near ubiquitous importance of global maize production. Deciphering this history is challenging, but new insight has been gained from examining 558 529 single nucleotide polymorphisms, expression data of 28 769 genes, and 662 traits collected from 368 diverse temperate and tropical maize inbred lines in this study. This is a new attempt to systematically exploit the mechanisms of the adaptation process in maize. Our results indicate that divergence between tropical and temperate lines apparently occurred 3400–6700 years ago. Seven hundred and one genomic selection signals and transcriptomic variants including 2700 differentially expressed individual genes and 389 rewired co-expression network genes were identified. These candidate signals were found to be functionally related to stress responses, and most were associated with directionally selected traits, which may have been an advantage under widely varying environmental conditions faced by maize as it was migrated away from its domestication center. Our study also clearly indicates that such stress adaptation could involve evolution of protein-coding sequences as well as transcriptome-level regulatory changes. The latter process may be a more flexible and dynamic way for maize to adapt to environmental changes along its short evolutionary history.

Key words: temperate adaptation, positive selection, regulatory evolution, *Zea mays*

INTRODUCTION

Maize (*Zea mays* ssp. mays) is essential to the global food supply, with current total maize grain production higher than any other crop (USDA FAS, 2013). Maize is also used as a model to investigate crop evolution and improvement (Doebley et al., 2006). It is thought to have been domesticated from teosinte (*Zea mays* ssp. *parviglumis*) about 9000–10 000 years ago in southwestern Mexico, which is a mid- to lowland tropical growing environment (Matsuoka et al., 2002; Van Heerwaarden et al., 2011). The remarkable conversion of a Mexican annual grass species into the top food, feed, and industrial crop in the world resulted from the spread of temperate maize over several thousand years from its tropical geographic origin to the north and east across North America and to the south across most of Latin America, eventually creating a maize distribution from ~40°S in Chile to ~45°N in Canada (Matsuoka et al., 2002). Centuries ago, maize cultivation expanded further to East Asia, Europe, and Africa, and the temperate–tropical division remains in all crop-growing continents today. When faced with widely varying temperate conditions in temperature, day length, and disease susceptibility, maize adapted remarkably well. One major goal of adaptation studies is to identify specific genomic changes contributing to advantageous phenotypic performance in varying environmental conditions.

In order to identify the genetic factors driving maize evolution, researchers have explored a number of methods to reveal footprints of selection within the genome (Chia et al., 2012; Hufford et al., 2012; Jiao et al., 2012). It is intriguing that these changes occurred within such a short evolutionary time frame. The importance of transcriptional regulation of rapid phenotypic evolution has been a central tenet of recent studies (Carroll, 2008; Ecker et al., 2012; Swanson-Wagner et al., 2012; Koenig et al., 2013). Genes with differential expression (DE) and altered expression networks could provide evidence of the contribution of transcriptome regulational changes to the adaptation process. RNA-seq allows cost-effective exploration of both sequence and transcriptional variation, particularly in large and repetitive sequence-rich genomes (Wang et al., 2009) such as maize.

Seed development, a critical process to both plant propagation and food supply, is a time in which DNA methylation and chromatin remodeling, and thus transcriptional patterns, are reshaped for the new generation (Ahmad et al., 2010; Van Zanten et al., 2011; Wollmann and Berger, 2012). Transcriptional variation may thus heavily influence seed-related traits via environmentally sensitive epigenetic control (Zhang and Ogas, 2009), which will be expressed as selectable variation throughout the lifetime of the plant (de Casas et al., 2012; Kapazoglou et al., 2013). Most maize genes are expressed in seed or embryos, many of which are not expressed again (Cho et al., 2002; Sekhon et al., 2011). Thus, the seed offers an important window into visualizing differences that may account for adaptation.

To study the nature of maize adaptation from tropical to temperate growing regions, a panel of 368 diverse maize inbred lines (Li et al., 2013) (Supplemental Table 1) was characterized. We combined RNA-seq of seeds (15 days after pollination; Fu et al., 2013) with data from the MaizeSNP50 BeadChip, resulting in over one million high-quality single nucleotide polymorphisms (SNPs) and expression quantitation from 28 769 genes, analyzed together with 662 phenotypic traits. These included morphological, agronomic, physiological, and metabolic traits, many of which are also known to be important in stress adaptation (Bohnert and Sheveleva, 1998; Bhargava and Sawant, 2013). This study seems to be the first systematic exploration of the mechanisms of the maize adaptation process, with the goal of answering several specific questions: Which phenotypic changes in temperate lines convey an advantage in novel environments? Which genomic regions were selected during the adaptation process? Which phenotypes do these regions likely affect? To what extent do regulatory changes contribute to evolution? What beneficial value do they provide in the relatively short evolutionary time frame suggested in this study? Although only one organ (the seed) was sequenced, such knowledge will position us with general understanding of the maize adaptation process and provide resources for developing breeding strategies to help corn producers cope with an increasingly erratic climate.

RESULTS

Population Level Differences between Temperate and Tropical Lines

The population-scaled recombination rates (ρ) in temperate and tropical lines were 1.078/kb and 2.644/kb, respectively. This is
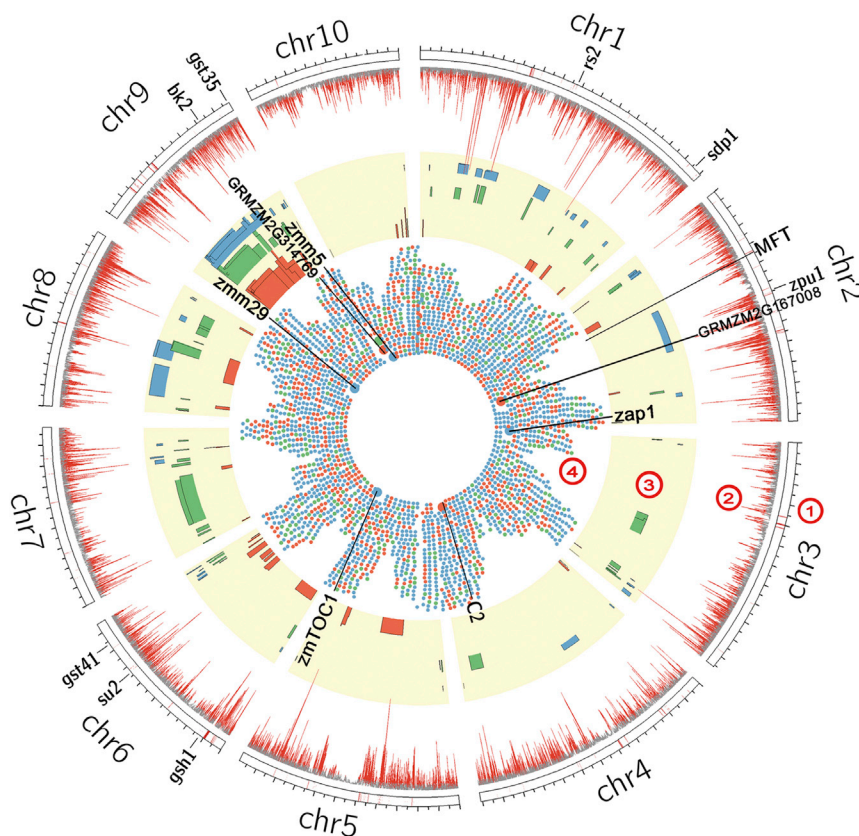
a reflection of different rates of linkage disequilibrium (LD) decay, which was much faster in the tropical lines at the whole genome level (Supplemental Figure 1). Recombination rate differences in temperate versus tropical lines was smaller than the decrease in ρ seen in a previous study by Hufford et al. (2012) in landraces compared with teosinte (59% vs 75%). A cross-population composite likelihood approach (Chen et al., 2010) (XP-CLR) was used to identify extreme allele frequency differentiation over linked regions when comparing temperate with tropical subpopulations. We identified 701 regions containing 1660 selected genes at the highest 10% of XP-CLR values (Figure 1 and Supplemental Tables 2 and 3), ranging in size from 10 kb to 2320 kb, with an average of 150.9 kb; this is shorter than the 322 kb average region associated with domestication (Hufford et al., 2012). The combined length of selected regions was 105.7 Mb, covering 5.2% of the genome.

Nucleotide diversity (π) in the selected features identified by XP-CLR in temperate versus tropical lines was 8.22 E−04 and 8.42 E−04, respectively, indicating a reduction of 2.4% in temperate lines. This decrease is less than the reduction of nucleotide diversity in selected features identified during domestication (17%) by Hufford et al. (2012). F_{ST} of selected features was 0.027, which is similar to 0.02 reported between landraces and improved lines and smaller when compared with 0.11 between teosinte and landraces, possibly due to the shorter time for adaptation from tropical to temperate than for domestication (Hufford et al., 2012). However, the features seem to be underestimated in this study, because most of our genotype was derived only from the genic region, while the whole genomic region in the study by Hufford et al. and intergenic regions were considered as more diverse.

The divergence time between temperate and tropical subpopulations is of interest and can be associated with the development of agriculture and the spread of human civilization in the Americas; however, archaeological information on this topic is incomplete and occasionally contradictory (Piperno and Pearsall, 1998; Staller and Thompson, 2002; Blake, 2006; Grobman et al., 2012). We proposed three models (detailed in the Materials and Methods section, Figure 2A–2C) to estimate the time of divergence, resulting in the estimation of 3400–6700 years BP. This time frame is supported by recent archaeological evidence (Haas et al., 2013) and implies that after domestication, maize cultivation rapidly expanded to temperate America (Figure 2D). The molecular evidence thus suggests that improvement and adaptation in maize may not have been sequential and discrete processes, but overlapped.

Genome-wide Selection Analysis and Functional Correspondence

The 701 selected regions were compared with 466 and 573 regions identified in domestication and improvement previously, respectively (Hufford et al., 2012; Supplemental Figure 2). Most adaptation regions did not overlap with domestication and improvement at all, indicating different genomic factors contributing to the phenotype changes during adaptation. Gene Ontology (GO) annotation of the top 571 candidate genes (see the Materials and Methods section) within the 701 selected regions reflected genes responding to stress, development,

Selected metabolic pathways responding to a changing environment are also involved in the adaptation process. Glutathione plays an important role in cellular processes under biotic stress (Dubreuil-Maurizi and Poinssot, 2012) and is one such example. Accordingly, *gst35* (glutathione *S*-transferase 35, GRMZM2G161891), *gst41* (glutathione *S*-transferase 41, GRMZM2G097989) and *gsh1* (γ-glutamylcysteine synthetase1, GRMZM2G111579), which influence glutathione metabolism, were all within our selected regions (Figure 1). Traits related to plant architecture and vegetative growth were also selected during adaptation, and their corresponding genes were found in our top XP-CLR hits, including *rs2* (rough sheath2, GRMZM2G403620), essential for normal leaf morphology (Phelps-Durr et al., 2005); *bk2* (brittle stalk2, GRMZM2G109326) affecting mechanical strength of maize by altering the composition and structure of secondary cell wall tissues (Ching et al., 2006), and *apt1* (aberrant pollen transmission1, GRMZM2G448687), affecting the elongation of root cortex cells and pollen tubes during temperature stress (Xu and Dooner, 2006) (Figure 1 and Supplemental Table 3).

Phenomic Changes Affecting Adaptation and Validation of Selected Regions

Adaptation involves selection of different ecotypes suited to different environments, leading to measurable phenotypic differences between environments. To test this assumption, we collected 662 traits, including agronomic, yield, seed quality, and seed metabolic traits (Supplemental Tables 5–7; Wen et al., 2014). The $Q_{ST} - F_{ST}$ method (Leinonen et al., 2013), by comparing complex partitioning of variation of quantitative traits and neutral molecular markers (see the Materials and Methods section for details), allowed us to distinguish whether the divergent traits are caused by directional selection ($Q_{ST} > F_{ST}$), genetic drift ($Q_{ST} \approx F_{ST}$), or stabilizing selection ($Q_{ST} < F_{ST}$) (Leinonen et al., 2013). One hundred and thirty traits displayed directional selection patterns ($Q_{ST} > F_{ST}$) across the populations and simultaneously showed significant ($P \leq 5.1E{-}05$) differences between the tropical and temperate subgroups (Supplemental Table 8) and were thus regarded as selected traits, which
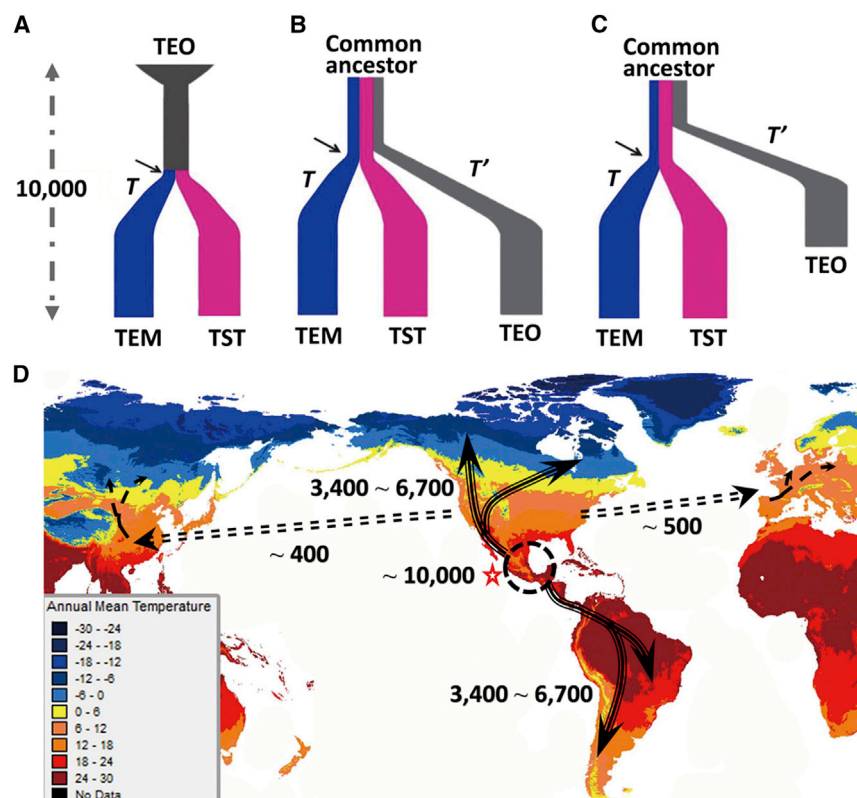
and metabolic processes (Supplemental Table 4). *MFT* (Mother of FT and TFL1, GRMZM2G059358) was identified as a strong candidate gene in adaptation (Figure 1). It encodes a maize MFT-like protein, which is involved in seed dormancy and germination, a complex adaptation process modulated through a negative feedback loop via abscisic acid (ABA) (Xi et al., 2010). *MFT* is also known to be involved in control of shoot meristem growth and flowering time (Yoo et al., 2004). Flowering time changes are critical in allowing maize to adapt to different environmental conditions, such as photoperiod and temperature. Another gene, GRMZM2G360455, an important locus affecting the photoperiod response based on the maize nested association mapping (NAM) population analysis (Buckler et al., 2009), and containing the CO/CO-Like/TOC1 conserved site (CCT), which always contributes to the circadian clock and flowering time (Robson et al., 2001; Griffiths et al., 2003; Cockram et al., 2012), was selected during maize adaptation (Figure 3A). Metabolic processes influencing nutritionally important traits such as starch content and oil concentration could likely be targets of selection not only during domestication and improvement but also adaptation, as different soil and climate conditions will influence the developmental stage of germinating maize. Some genes associated with nutritional traits were selected during the adaptation process in our study (Figure 1), such as *su2* (GRMZM2G348551), *zpu1* (pullulanase-type starch debranching enzyme1, GRMZM2G158043), and *sdp1* (GRMZM2G087612); these have also been identified as selection targets during domestication and improvement (Beatty et al., 1999; Zhang et al., 2004; Eastmond, 2006).

Figure 2. Maize Dispersion Map and Divergence Time Estimations.

(A) Divergence time (T) was estimated between temperate (TEM) and tropical/subtropical (TST) maize, using extant teosinte (TEO) lines as the ancestor of maize.

(B) Divergence time (T) was estimated using a model supposing a common ancestor to TEM, TST, and TEO and equal selection pressure on extant TEO lines (leading to T' for teosinte) as on TST and TEM.

(C) Divergence time was estimated between **(A)** and **(B)**, with the more likely assumption that TEO lines have experienced a lower level of selection pressure (and smaller T') than the intensity faced by TEM and TST during the adaptation process.

(D) Maize dispersion map showing environmental difference in annual mean temperature. Arrows show possible dispersion routes and the numbers beside the routes indicate the likely time (in years before present) when the dispersion happened, and the circle (beside a silver star) is the center of maize domestication (Hufford et al., 2012). Even although there are no inbred lines from South America, the dispersion route to South America was inferred from a previous study (Wallace et al., 2014).

were likely to contribute to improved phenotypic performance in temperate conditions.

To test if genetic variation within selected regions contributed to these phenotypic changes, a genome-wide association study (GWAS) was performed on the 130 divergent traits with 24 595 SNPs (minor allele frequency [MAF] $\geq$ 5%; only 4.4% of the total) from the selected regions. In total, 345 selected regions (49.22%) were associated ($P < 4.06\text{E}-05$) with 100 (or 76.92%) of the 130 traits, including three agronomic (days to silking [DTS], cob color, and kernel color), one amino acid (Ala), and 96 metabolic traits (Supplemental Table 9). The genes identified here undoubtedly represent only a fraction of the total genes selected during adaptation, since many target traits were not measured in this study.

Quantitative trait locus (QTL) mapping was used to confirm the phenotypic effect of the selected regions. Flowering time changes allow maize to adapt to temperate environmental conditions such as photoperiod and temperature. Three recombinant inbred line (RIL) populations generated by crossing temperate lines with tropical/subtropical lines were used to map flowering time (Supplemental Table 10). The identified QTLs are reported in Supplemental Table 7, and many (113, or 62%) overlapped significantly ($P < 2.2\text{E}-16$) with the identified selected regions (Figure 1). For example, gene GRMZM2G360455, encoding a CCT domain-containing protein, was located in the QTL interval for DTS under short day (tropical) but not under long day (temperate) conditions, and also, a candidate response to flowering time of the maize NAM population (Buckler et al., 2009) was identified within a selected region (Figure 3A–3C). Further study of this gene may
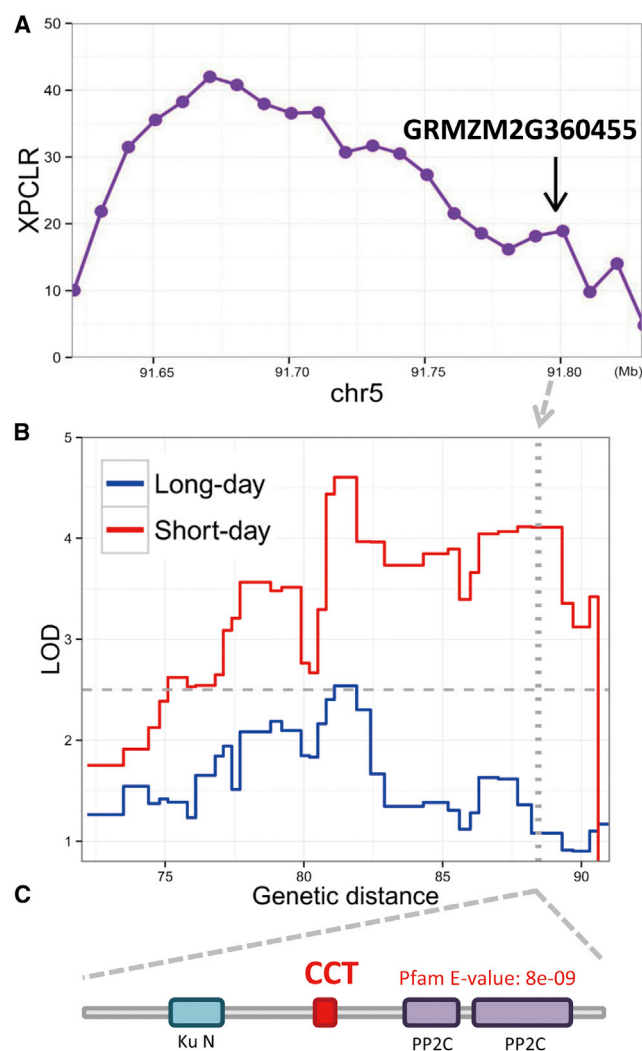
provide a better understanding of maize flowering time and adaptation.

Significance of Transcriptional Regulation to Adaptation

Changes in gene regulation, affecting gene expression level but not gene structure, are fundamental to the evolution of morphological and developmental diversity (Carroll, 2008; Swanson-Wagner et al., 2012). The present data sets provide an excellent resource for investigating the contribution of transcriptome regulation to the adaptation process. The coefficient of variation, a normalized measure of dispersion of a probability distribution for expression quantification, was similar between tropical and temperate germplasm when considering gene expression of all genes (Figure 4A), suggesting that most inbreds were probably at the same developmental stage and that no dramatic transcriptome-wide changes were observed between tropical and temperate lines. This agrees with a previous study (Swanson-Wagner et al., 2012), indicating that overall changes in expression did not happen during domestication and post domestication. To study specific transcriptome signal response of individual genes or groups of genes contributing to the adaptation process, Q_{ST}–F_{ST} of differentially expressed genes were compared, including single genes differentially expressed under different conditions or in different samples (DE) and genes with altered expression conservation (AEC) (Swanson-Wagner et al., 2012), representing the rewired co-expression of a gene network.

DE

Among 28 769 expressed genes, 2700 (9.4% of the total) showed significant DE (posterior probability >0.9999) based on a regularized *t*-test method; all exceeded neutral expectation

Figure 3. Example of Validation of the Function of Genes within Selected Genomic Regions Using QTL Mapping.

(A) Gene GRMZM2G360455 was in a selected region on chromosome 5.
(B) The region that contained GRMZM2G360455 covered a QTL for DTS.
(C) Annotation of GRMZM2G360455 disclosed a CCT domain in this gene, which is related to flowering time.

(Supplemental Table 11). Comparing between temperate and tropical lines in a publicly available database with the same quantitative measurement and DE analysis process, we found that a major part of these DE genes were also expressed differently in other tissues, including the shoot apex (Li et al., 2012) (P = 1.88E−38) and seeding L3 leaf (Eichten et al., 2013) (P = 1.44E−15; Supplemental Figure 3). This is highly suggestive that most DE genes identified here were not caused by random environmental and developmental variation and that many of the candidate DE genes continued to be important to adaptive differences later in the development of the mature plant. With more relaxed posterior probability (>0.999), 14.4% of the total genes showed expression differences and most were still likely to have been caused by directional selection (Supplemental Figure 4).

There were 871 up- and 1829 downregulated genes in temperate versus tropical lines (Figure 4B and Supplemental Table 11).
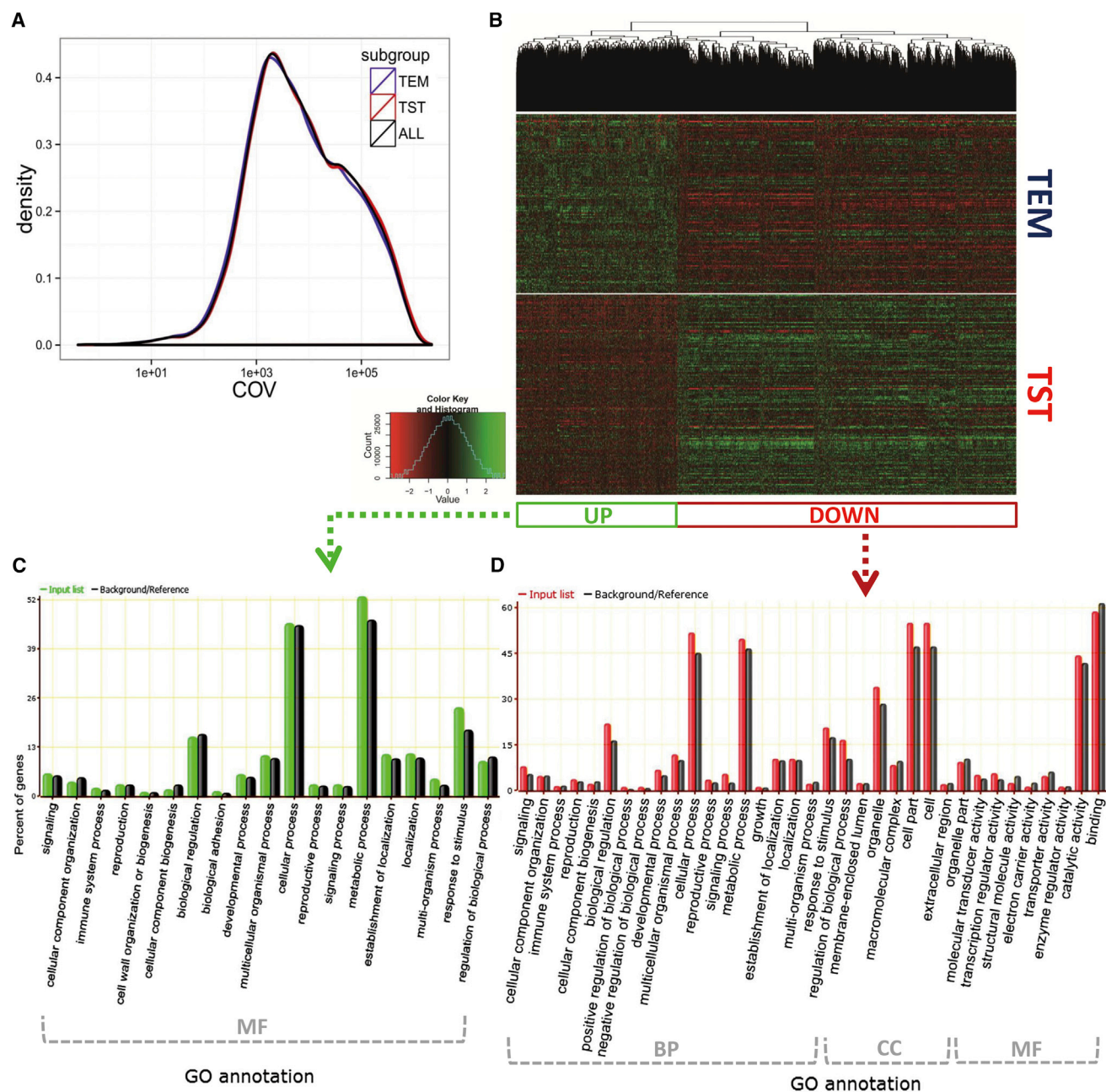
These DE candidates tend to be regulated by distant eQTLs (P = 1.74E−4; eQTL data from Fu et al., 2013), and especially for upregulated (P = 9.67E−12) but not for downregulated (P = 0.83) genes in temperate maize. With local (cis-) regulation, expression differences are caused by one regulatory region near each expressed gene, while more distant (trans-) regulation often sees the expression of whole groups of genes regulated by a single genetic factor, causing potential widespread pleiotropic effects. Thus, trans-regulatory mutations seem to be better suited to the changes of complex phenotypes, which are governed by the coordinated expression patterns of multiple genes and a single regulator. While some previous studies have emphasized the key role for cis-regulatory evolution (Wray, 2007; Stern and Orgogozo, 2009), our results suggest that trans-regulatory variation could contribute more commonly to adaptive phenotypic divergence.

GO analysis (Figure 4C and 4D) of DE genes showed an enrichment of upregulated genes involving molecular function, especially in catalytic activity, oxidoreductase activity, endopeptidase inhibitor activity, and transferase activity (Supplemental Figure 5). Endopeptidase levels are influenced by stress in plants (Richen et al., 2003; Antão and Malcata, 2005) and animals (Karlsson et al., 2006), but clear mechanisms are still unknown. The GO analyses of downregulated genes in temperate lines revealed enrichment of genes involved in processes such as response to stimuli, metabolism, and regulation (Supplemental Figure 5). The same GO analyses also uncovered genes involved in cellular components and molecular functions; one downregulated protein serine/threonine phosphatase involved in both aspects was particularly significant and is associated with biotic and abiotic stress (Antão and Malcata, 2005; Supplemental Figure 5).

The MADS-box family of transcription factors is important in the evolution of plant architecture and angiosperm inflorescence development and is frequently identified as targeted regions of selection during the domestication of maize (Zhao et al., 2011). Three MADS-box genes (zmm5 [GRMZM2G148693], zmm29 [GRMZM2G152862] and zap1 [GRMZM2G171365]) were all downregulated in temperate lines (Supplemental Figure 6). These genes belong to the MICK[c] class of the MADS-box gene family and the TM3, GLO, and SQUA subfamilies, respectively, and are involved in growth and flower homeotic function (Münster et al., 2002). The circadian clock is vital in flowering time networks, which consist of three negative feedback loops. Gene TOC1b (GRMZM2G148453) is located in the central loop (Kolmos et al., 2008) and showed downregulation in temperate lines in the current study (Supplemental Figure 6). While some classical genes were also identified as DE even during domestication and improvement processes (Supplemental Figure 6), only 129 genes were identified both in DE and genome selection in our analysis (Supplemental Figure 7).

DE-GWAS

To seek association between DE genes and the divergent traits, transformed expression values of DE genes were analyzed via DE-GWAS (see the Materials and Methods section) with the 130 traits that had been found to be significantly different between tropical and temperate lines. Two hundred and forty-five DE genes (9.07% of the total) were associated (P < 1E−3) with 101 traits, including 75 traits overlapping with those detected by trait-SNP GWAS (Supplemental Table 12). The 101 DE

Figure 4. Transcriptional DE Analysis.
(A) Density plots for the coefficient of variance (COV) for gene expression levels in all lines (black), temperate lines (blue), and tropical lines (red).
(B) Genes with significantly different expression were used for hierarchical clustering.
(C) Enrichment analysis of GO annotation within upregulated genes in temperate lines relative to tropical lines.
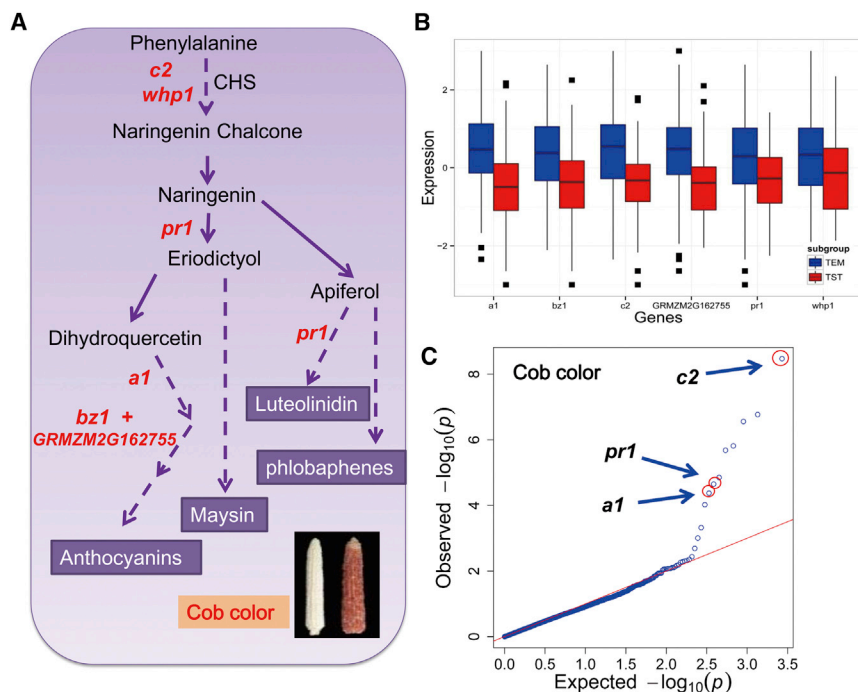**(D)** Enrichment analysis of GO annotation of downregulated genes in temperate lines relative to tropical lines. Within the specific GO terms in **(C)** and **(D)**, MF represents genes of molecular function, CC represents genes with a cellular component, and BP are genes associated with biological processes.

associated traits included 6 agronomic (cob color, DTS, days to tasseling (DTT), days to pollen shedding, leaf number, and kernel color), 5 amino acid (Ala, Arg, Asp, Lys, Gly) and 90 metabolic traits (Supplemental Table 12). Seven of the genes identified as associated in the genomic sequence analysis overlapped with genes identified in the DE-GWAS analysis.

A series of DE genes (*c2*, *pr1*, *a1*, *bz1*, *whp1*) were detected that influence cob color via the flavonoid biosynthetic pathway

(Figure 5; Sharma et al., 2012). Cob color segregated only in the temperate group and is known to be affected by chalcone synthase (CHS) and maysin synthesis, which are thought to be major contributors to corn worm resistance (McMullen et al., 2009; Sharma et al., 2012). As can be seen in Figure 5, *c2* (colorless2, GRMZM2G422750) is one of the main genes of the CHS pathway and is located upstream of maysin synthesis (Sharma et al., 2012). *pr1* (purple aleurone1 or red aleurone1, GRMZM2G025832) is located downstream of *c2*, encodes a

Figure 5. Cob Color Was Influenced by the Flavonoid Biosynthetic Pathway in Maize.

(A) A simplified flavonoid biosynthetic pathway. Genes in red were found to be associated with cob color in this study.

(B) Genes involved in the flavonoid pathway showed significantly ($P < 5.06E-06$ for all six candidates) DE between temperate and tropical groups.

(C) Three DE genes (*c2* [GRMZM2G422750], *pr1* [GRMZM2G025832], *a1* [GRMZM2G026930]) had significant association with cob color.

CYP450-dependent flavonoid 3′-hydroxylase required for synthesis anthocyanin, and is involved in naringenin chalcone catabolism (Sharma et al., 2012). *a1* (anthocyaninless1, GRMZM2G026930), located downstream of *c2* and *pr1*, is involved in the production of anthocyanins (Sharma et al., 2012). *c2* and *a1* were significantly associated with flavonoid catabolism (Supplemental Table 12). Kernel color ranked from light to dark (Supplemental Figure 8A) displayed different distributions between tropical and temperate subgroups (Supplemental Figure 8B) and was slightly negatively correlated with cob color (Supplemental Figure 8C; $P = 1.31E-05$). Kernel color was also affected by DE genes within the CHS pathway, and different association patterns were observed between tropical and temperate lines by DE-GWAS analysis (Supplemental Figure 8D).

AEC

While DE can identify individual differentially expressed genes, AEC (Swanson-Wagner et al., 2012) reflects the relationship of genes with their co-expressed group. In total, 389 genes showed the strongest AEC patterns (expression conservation score >2.5 SD; Supplemental Table 13). Further analysis indicated that the average number of genes significantly co-expressed with AEC candidates (as measured by an absolute Pearson correlation coefficient ≥0.3) in temperate lines was higher (259 genes) than in tropical lines (147 genes; Supplemental Figure 9). This indicates that the changes faced in temperate environments may have enhanced interactions between genes in certain pathways. Among the stronger relationships, the number of genes with negative correlations (Pearson correlations coefficient ≤0) was higher in temperate germplasm as well (24.4% in temperate lines vs 3.2% in tropical lines). Few genes were co-expressed with the same candidate AEC gene in both temperate and tropical lines (Supplemental Figure 9), which suggests a rewiring of the regulation networks in the temperate subgroup during adaptation. However, the rewiring appears to have been less

dramatic during adaptation, since fewer AEC genes were identified for adaptation than for domestication (Swanson-Wagner et al., 2012). In contrast, the number of differentially expressed individual genes found in the adaptation analysis was higher than for domestication (Swanson-Wagner et al., 2012).

As sessile organisms, plants have evolved to integrate endogenous and external information and use signal transduction processes to allow growth plasticity and survive and thrive in their environments. Essential to this plant–environment interaction are plant hormones, including auxins, ethylene, ABA, and brassinosteroids, which play a key role in plant growth, development, defenses, and stress tolerance (Wolters and Jürgens, 2009). Plant hormones modulate gene expression by controlling either the abundance of transcriptional factors or repressors or their activities through post-translational modifications (Dharmasiri et al., 2013). Hormonal crosstalk and interaction with other plant compounds and environmental factors rely on complicated signaling networks. These networks generally intersect in central nodes. Six candidate genes related to plant hormones were observed in our AEC analysis and may act as central nodes in plant hormonal production. Three (GRMZM2G078480, GRMZM5G860241, and GRMZM2G086773) belong to the brassinosteroid biosynthesis pathway: one (GRMZM5G864847) was an auxin-responsive Aux/IAA family member and two (GRMZM2G026131, GRMZM2G390385) were in the pathway producing ethylene biosynthesis from methionine (Supplemental Figure 10). These candidates, along with their co-expressed genes, showed very different networks in temperate versus tropical germplasm (Supplemental Figure 10). This could provide helpful clues to a deeper understanding of the complex relationship between hormones and their contributions to maize adaptation.

Many studies have identified the function of F-box receptors in hormone controlling and signaling (Dharmasiri et al., 2005; Kepinski and Leyser, 2005; Koops et al., 2011; Shen et al., 2012). AEC analysis also identified an F-box protein (GRMZM2G031958; Jia et al., 2013) displaying different regulation patterns (Supplemental Figure 11), with a series of 462 and 85 highly co-expressed genes observed in temperate and tropical lines, respectively. Beyond the remarkable difference in network size, several co-expressed enzymes, including

cytochrome-c reductases, NADH-ubiquinone oxidoreductases, peroxidases, and amidase, and some genes that functioned in ABA and indole-3-acetic acid biosynthesis and gluconeogenesis exist only within temperate lines. These results are consistent with the earlier studies that F-box proteins could play a regulatory role in glucose-induced seed germination by targeting ABA synthesis (Song et al., 2012). F-box proteins also play critical roles in seed development, grain filling, and response to abiotic stress (Jain et al., 2007) in crop plants, and co-expressed (with our F-box candidate) genes involved in cytokinins degradation, cellulose biosynthesis, and flavonoid and flavonol biosynthesis pathways were observed in the temperate subgroup but not in the tropical. This F-box gene was also highly co-expressed with an ethylene-responsive factor-like protein (GRMZM2G169382), an abscisic stress protein homolog (GRMZM2G044132), a SAUR37-auxin-responsive family member (GRMZM2G045243), another two F-box domain-containing proteins (GRMZM2G116603 and GRMZM2G071705), and several transcriptional factors (zmm29, GRMZM2G152862; ethylene-responsive transcription factors, GRMZM2G474326 and GRMZM2G169382; WRKY1, GRMZM2G018487; WRKY25, GRMZM2G148561; WOX2B, GRMZM2G339751; and more with putative regulation of transcription functions). In addition, the teosinte glume architecture 1 (tga1) protein and an Early Flowering 4 (EF4)-like protein simultaneously showed strong relationships with the target F-box gene. These two proteins have been hypothesized to contribute to maize adaptation to temperate climates in past studies (Khanna et al., 2003; Ducrocq et al., 2008). The identification of the genes encoding these proteins made the network more complex, and further studies are needed. GO analysis of all the co-expressed genes in temperate lines revealed an abundance of genes that respond to temperature stimulus and abiotic and other stress (Supplemental Table 14), which were undoubtedly required during maize adaptation to temperate environments.

DISCUSSION

The process of adaptation was as important a driving factor as domestication in the creation of a geographically diverse crop, allowing maize to spread into a wide range of environments around the world. The impacts of domestication and improvement on the genome and transcriptome have begun to be studied (Hufford et al., 2012; Swanson-Wagner et al., 2012), while adaptation has not yet been as systematically analyzed. In this study, large sets of genomic, transcriptomic, and phenomic data were used to analyze the mechanisms of morphological evolution leading to adaptation. Plant response to environmental change, especially stress, may have been the key initial step towards adaptation to more extreme latitudes (Figure 2D). A variety of mechanisms could have contributed to stress response during the maize life cycle, including changes in seed dormancy, germination, plant architecture, flowering time, and optimal utilization of resources (nitrogen, water, etc). In the face of new biotic stresses, a series of resistance mechanisms can also evolve at both the genomic and transcriptomic levels. Traits allowing the plant to successfully respond to stress are precisely those of interest to farmers and plant breeders who have achieved improvement of temperate lines by the selection of such stress tolerance (Tollenaar and Wu, 1999).

It has been suggested that studies scanning for positive selection may incur high false-positive rates and can be misleading (Pavlidis et al., 2012). In this study, we use two additional and reliable methods, GWAS and QTL mapping, to provide stronger evidence for the selected regions and a link to the contributory traits. We also identified transcriptomic variants contributing to the adaptation process, including differentially expressed individual genes and evidence for rewiring of co-expression networks. These candidate genes and regions were found to be functionally related to stress response and most were associated with the directionally selected traits. While this study focused on seed transcriptome, the seed-expressed genes and phenotypes provide the first steps towards a systematic study of the adaptation process and inform our understanding of the extent to which transcriptome variation influences the environmental adaptation process.

It has recently been reported that human adaptation is driven primarily by gene expression changes (Fraser, 2013). The present study reveals that transcriptome regulation was also prevalent in maize adaptation. Plants live in a dynamic environment, but selection on genomic variation may be too slow to cause the changes that allow the plant to adapt to a rapidly changing environment (Chen, 2007). Selection on protein-coding sequence variation is risky, as most mutations are harmful or lethal to the organism, and even changes that are beneficial to some cells or under some conditions may be harmful to other tissues or under other conditions (Ecker et al., 2012). Sufficient protein-coding sequence changes at the genomic level would be unlikely to respond to environmental changes experienced in one or a few generations; however, rapid changes in transcriptome regulation can occur quickly and lead to a rapid phenotypic differentiation (Chen, 2007). Transcriptome changes are resource economical and are frequently associated with temporally and spatially related gene expression patterns, the effects of which can be limited to specific cells (Carroll, 2008; Ecker et al., 2012). Our models, indicating a separation time between temperate and tropical maize of 3400–6700 years BP, suggest that a great number of changes took place during a few thousand years. The differences in transcription of individual genes and correlated suites of genes between temperate and tropical maize can be explained by this hypothesis. Some studies (Ptashne, 2007; Cortessis et al., 2012) have suggested that epigenetic regulation was the main genetic driver causing regulatory evolution and that epigenetic modification could also lead to increased mutation rate (Rideout et al., 1990; Schuster-Böckler and Lehner, 2012). Further study and more direct evidence will be needed to better understand the interplay between epigenetic and genetic processes under selection and provide new insights, and possibly new mechanisms, for practical plant improvement.

Although our comprehensive study of genomic, transcriptomic, and phenomic variation sheds new light on the process of adaptation in modern maize, much is left to be uncovered. In particular, changes due to adaptation of temperate maize are inferred in the current study by comparing temperate maize with tropical maize grown in northern temperate growing areas. Although similar divergence and changes may have happened as maize migrated to far southern temperate regions, no South American maize was studied here, and thus, this conclusion remains to be confirmed.

In future studies, expression data from more tissues and genotypes need be included and studied under more environmental conditions to allow a finer dissection of genes and mechanisms involved in adaptation. It also should be kept in mind that all the genotypic variation was identified by comparison with the temperate maize reference genome (B73). If a bias was introduced by this method of polymorphism identification, it was not considered in this study. Novel assembly strategies taking into account the variation from more lines could reduce this bias (if present), and exploration of more variant types, such as presence/absence variation, should also be considered in future studies of the adaptation process. Detailed studies on changes in the transcriptome and, in particular, the role of epigenetics could lead to a clearer understanding of adaptation, possibly leading in turn to more innovative techniques to allow plant breeders to apply native trait variation to maize improvement.

MATERIALS AND METHODS

Maize Inbred Lines and Collection of Genotypic, Phenotypic, and Gene Expression Data

The 368 maize inbred lines included in this study form a global collection, including representative temperate and tropical/subtropical inbred lines listed in Supplemental Table 1, and additional information about the lines can be found in a previous study (Yang et al., 2011).

The poly(A) transcriptome collected from kernels at 15 days after pollination from all 368 lines was sequenced using 90-bp paired-end Illumina sequencing with libraries of 200-bp insert sizes, and 25.8 billion high-quality reads were obtained after filtering out reads with low sequencing quality. A total of 1.03 million high-quality SNPs with a missing data rate less than 60% were used for imputation of missing genotypes. Three pairs of biological replicates (SK, Han21, and Ye478) were used to evaluate the reproducibility of genotyping by RNA-seq. The concordance rates were greater than 99% between each pair of replicates (Fu et al., 2013), indicating that the sequencing and SNP calling procedures were reproducible. In addition, all lines were genotyped via the Illumina MaizeSNP50 array. SNPs generated by RNA-seq also met high concordance rates with the genotypes determined by the MaizeSNP50 BeadChip (Fu et al., 2013). Both RNA-seq and MaizeSNP50 data sets were merged to obtain a total of more than 1.06 million SNPs; a final data set of 558 529 SNPs with a MAF >5% was produced, and more detailed information can be found in a previous study (Fu et al., 2013).

To quantify the expression of known genes with annotation in the B73 reference genome (filtered-gene set, version 2, release 5b), 28 769 genes corresponding to mapped sequence reads in more than 50% of the inbred lines were compiled, and each gene averaged more than 1.5 K reads. Read counts for each gene were calculated and scaled according to RPKM (reads per kilobase of exon model per million mapped reads). After RPKM normalization, all genes with a median expression level greater than zero for each sample were included, and the overall distribution of expression levels for each gene was normalized using a normal quantile transformation (Fu et al., 2013). The same three pairs of biological replicates were shown to share most compiled genes (average 95.71%) with high concordance (average person's r = 0.87, (Supplemental Figure 12) in expression quantification between each pair of replicates. More details on library construction, sequencing, SNP detection, genotype imputation, positive control of SNP accuracy, and quantile normalization of expression are described in Fu et al. (2013).

To obtain agronomic traits (reported in Supplemental Table 5; Yang et al., 2014), all inbred lines were planted with randomized complete experimental design by single replication in 2010 in four locations (Honghe autonomous prefecture, Yunan province; Sanya, Hainan

province; Wuhan, Hubei province; Ya'an, Sichuan province) and 2011 in three locations (Chongqing; Hebi, Henan province; Nanning, Guangxi province). The seven different locations ranged from 18 to 35 degrees north in latitude and from 102 to 114 degrees east in longitude. Kernel color was measured in five additional trials over 2 years (Sanya, Hainan province in 2011; Honghe autonomous prefecture, Yunan province in 2012; Chongqing in 2012; Wuhan, Hubei province in 2012; Hebi, Henan province in 2012). All agronomic traits were measured and the best linear unbiased predictor (BLUP) values from different environments and years were used for the final analysis (Supplemental Table 2). Maize kernels from each entry in the panel planted in Chongqing in 2011 were collected to quantify the amino acid content (Supplemental Table 6), and samples from the panel planted in Yunnan (2011) and Hainan (2010) were harvested to measure metabolic traits (Wen et al., 2014). Only the metabolites measured in at least two of the three experiments and showing high correlation (P < 0.05) in the two experiments were retained to calculate BLUP values for the next analysis (Supplemental Table 7).

Population Structure of 368 Inbred Lines

The STRUCTURE software package (Pritchard et al., 2000) was used to analyze the population structure, and the EIGENSOFT analysis package (Patterson et al., 2006; Price et al., 2006) was used to run a principal component analysis of the panel used in the present study (Supplemental Figure 13). Considering the computation time, an SNP marker subset was used for inferring the population structure; the subset of 14 685 SNPs was created by removing adjacent SNPs within 50 kbp intervals. All lines were divided into three subpopulations corresponding to stiff stalk (SS), non-stiff stalk (NSS) and tropical and subtropical (TST) clusters by STRUCTURE, using a probability of inclusion into each cluster of greater than 0.65 (Yan et al., 2009). In the final analysis, there were 133 temperate (103 NSS + 30 SS) lines, 149 tropical (TST) lines, and 86 mixed lines (Supplemental Figure 13 and Supplemental Table 1). The mixed were excluded from further analyses to allow a clearer comparison between tropical and temperate features. The population structure was similar to the analysis of the same lines reported previously using 1536 SNPs (Yang et al., 2011).

Measuring the Genomic Changes Occurring during the Maize Adaptation Process

To evaluate changes in the maize genome due to adaption, population genetics statistics, including π (Tajima, 1983) and F_{ST} (Nei, 1977), were calculated within the differently adapted maize groups. F_{ST} was estimated as follows:

$$F_{ST} = \frac{(H_T - H_S)}{H_T}$$

$$H_S = 1 - \omega_{TEM} \times \left(freqA_{TEM}^2 + freqB_{TEM}^2\right) - \omega_{TST} \times \left(freqA_{TST}^2 + freqB_{TST}^2\right)$$

$$H_T = 1 - (\omega_{TEM} \times freqA_{TEM} + \omega_{TST} \times freqA_{TST})^2 - (\omega_{TEM} \times freqB_{TEM} + \omega_{TST} \times freqB_{TST})^2$$

where H_S refers to heterozygosity within subpopulations, and H_T refers to heterozygosity in the overall population. The variable ω refers to proportion of subpopulation based on size, and freq$A$/freq$B$ is the frequency of allele A or allele B in each subpopulation. The Synbreed package (Wimmer et al., 2012) of the R-project statistical package (http://www. R-project.org) was used to compute the LD coefficient, r^2. The ggplot2 package of R (Wickham, 2009) was used to plot LD decay, as well as all visualizations in this study (except as noted, below).

To identify regions of the maize genome that have undergone selection during the process of adaption from tropical to temperate climates, XP-CLR (Chen et al., 2010) was used. The following parameters were applied to implement the XP-CLR test: the size of the window was 0.005 Morgan; the maximum number of controlled SNPs within a window

was 100; the spacing between two grid points was 1 kb; and a corrLevel value of 0.7 was used as a down-weighted criterion in the weighted composite likelihood ratio test. Adjacent 10-kb windows from the top 20% of the XP-CLR results were merged into larger regions, according to Hufford et al. (2012), and only one window lower than 20% was retained for each region. Regions in the highest 10% of the mean region-wise XP-CLR values were regarded as having undergone selection. Gene sequences closest to the maximum XP-CLR value were designated as the most likely candidate genes for selection, while others within each selected region were considered as possible selected genes. The linkage map used in the XP-CLR analysis was constructed using an RIL population generated from the cross of B73×BY804 with 197 individuals described previously (Chander et al., 2008), and the maize SNP50 chip (Ganal et al., 2011) was used to re-genotype the RIL population with 15 285 polymorphic SNPs. The distance between unmapped SNPs was estimated based on the constructed linkage map and B73 reference genome.

Estimating the Relative Divergence Time between Temperate and Tropical Lines

Assuming a time scale of teosinte/maize divergence of about 10 000 years, an F_{ST}-based approach (Schlebusch et al., 2012) was used to estimate a relative divergence time between temperate and tropical subpopulations, under the assumptions of no genetic drift, no change in effective population size, and equal generation times in each lineage. Although violations of these assumptions are probable and may reduce the accuracy of the estimated divergence times, the estimated divergence times will still be useful for the purposes of this study and in the general study of maize evolution. We combined our SNP data with the maize HapMapII (Chia et al., 2012) data and retained all the SNPs from teosinte (Chia et al., 2012) and maize with the same loci, consistent alleles, and missing ratio of alleles less than 20% to calculate F_{ST} between temperate maize (TEM) and teosinte (TEO, 0.0668), and between TEM and tropical/subtropical maize (TST, 0.0453). Divergence time (T, measured in units of 2Ne generations, where Ne is the effective number of diploid individuals) was calculated using F_{ST} according to the following formula (Schlebusch et al., 2012):

$$T = -\log(1 - F_{ST})$$

Most of the SNP genotypes were located within expressed sequences, but the results should not be affected by this, assuming similar biases between the two comparisons and sufficient numbers of markers to smooth out unequal biases due to potential unequal selection pressures on some loci. Different models were proposed to improve estimation of the relative divergence time. Assuming that the teosinte lines from which we extracted SNPs were indeed the primitive ancestors (or contained the same sequence diversity within them as the actual ancestors), then $T_{TEM-TEO}$ = 10 000 years; (Figure 2A) and we can calculate the divergence time between TEM and TST using the following formula:

$$\frac{T_{TEM-TST}}{T_{TEM-TEO}} = -\frac{\log(1 - Fst_{TEM-TST})}{-\log(1 - Fst_{TEM-TEO})}$$

This resulted in a divergence time of $T_{TEM-TST}$ = 6700 years BP. Assuming further that the teosinte lines have undergone a similar selection pressure (Figure 2B), the analogical formula can be used to calculate $T_{TEM-TST}$ divergence time as 3400 years BP. However, because it is more likely that the teosinte lines have experienced a selection level that is not as strong as that of the adaptation process (Figure 2C), the true divergence time of adaptation would fall between the two estimated times. Bioclimatic variables data from WORLDCLIM database (Hijmans et al., 2005) and the DIVA-GIS (Hijmans et al., 2012) software were used to map the stress (annual mean temperature) faced by maize during the adaptation process (Figure 2D).

Analysis of Directional Selection of Phenomic Divergence

$Q_{ST} - F_{ST}$ comparison provides us with a method to distinguish population differentiation of complex polygenic traits as having been caused by natural selection or by genetic drift (Leinonen et al., 2013). Q_{ST} was estimated for all phenomic traits, including expression traits, as follows:

$$Q_{ST} = \frac{\sigma_{GB}^2}{\left(\sigma_{GB}^2 + 2\sigma_{GW}^2\right)}$$

where σ_{GB}^2 refers to the genetic component of variance among subpopulations, and σ_{GW}^2 is the average component of variance within each subpopulation.

To accurately estimate the distribution of mean F_{ST} among tropical and temperate subpopulations, 10 000 SNPs were chosen randomly from the entire SNP data set and the calculation was repeated 1000 times (Supplemental Figure 14). By applying a strict outlier definition, we used a 99% confidence interval (0.025–0.0273) for the F_{ST} distribution, ensuring a more correct comparison of $Q_{ST} - F_{ST}$. A $Q_{ST} > F_{ST}$ value was regarded as proof of trait divergence outstripping the expectation of a neutral state and thus of a strong directional selection signal (Leinonen et al., 2013).

DE Analysis

Cyber-T (Baldi and Long, 2001), a regularized t-test method that also contains statistical inferences on experiment-wide false-positive and false-negative levels based on the modeling of P value distributions, was used on normalized expression data (Fu et al., 2013) for the identification of statistically significant differentially expressed genes. Posterior probability of DE ≥ 0.9999 was determined to identify DE genes in our study.

Characterization of Genes Displaying AEC

To identify which genes show AEC, a statistic that reflects the co-expression of genes in a gene network (Swanson-Wagner et al., 2012), the expression data were divided into two matrices (E^{TST} and E^{TEM}) based on adaptation (tropical and temperate), and a co-expression network was created for each matrix (R_{ij}^{TST} and R_{ij}^{TEM}). The hmisc package in R (Harrell, 2012) was used to calculate the Pearson correlation coefficient between each pair of gene expression values within each subset (TEM or TST). Hmisc is an efficient algorithm for calculations in very large data sets, and is calculated as:

$$R_{ij}^{TST} = PCC\left(E_i^{TST}, E_j^{TST}\right)$$

$$R_{ij}^{TEM} = PCC\left(E_i^{TEM}, E_j^{TEM}\right)$$

for $i,j = 1,…,28\ 769$. Thus, R_{ij}^{TST} and R_{ij}^{TEM} are square matrices with the same dimensions (28 769 × 28 769). Each value in the two matrices represents an edge weight in the co-expression network and is the measured similarity between expression profiles of paired genes. Although the two matrices have identical dimensions, the distribution of values in each matrix may differ because of unequal sample sizes. Thus, to compare the two co-expression networks more accurately, the distributions were normalized by subtracting the mean and dividing by the SD to obtain a standard normal distribution. An expression conservation (EC) score was calculated as the Pearson correlation coefficient (PCC) between gene profiles in the two co-expression networks as described by Swanson-Wagner et al. (2012):

$$EC = PCC\left(E_i^{TEM}, E_i^{TST}\right)$$

where E_i^{TEM} and E_i^{TST} were represented by the ith rows in the temperate and tropical co-expression matrices, respectively. AEC genes were selected using a z score (Swanson-Wagner et al., 2012) to calculate each EC value as follows:

$$z = \frac{EC - \mu}{\sigma}$$

where μ and σ were the mean and SD of all the EC scores of the gene. The z score cutoff of AEC values of genes was set at ≤ -2.5. The software

Circos (Krzywinski et al., 2009) was used for visualizing the results in a circular layout.

Candidate Gene Annotation and GO Enrichment Analysis

To more fully explore the functions of candidate genes, the annotation resources of maizeGDB (Lawrence et al., 2008) and the InterPro (Zdobnov and Apweiler, 2001) database were integrated into the analyses. GO enrichment analysis was maintained by AgriGO (Du et al., 2010) with the Fisher statistical test method ($P \leq 0.05$) and Yekutieli multitest adjustment method (False Discovery Rate ≤ 0.05). The GOSlimViewer of AgBase (McCarthy et al., 2006) was used to implement GO slim analysis, and the updated GO items of the maize genome were downloaded from Ensembl BioMart (Kinsella et al., 2011) on April 4, 2013.

Association and QTL Mapping Analysis

SNPs within regions that have experienced selection according to the XP-CLR approach (24 595 in total) were used in an association analysis with the 111 different traits using the software package GAPIT (Lipka et al., 2012) and a compressed mixed linear model. The cutoff for significance for associations was set at $1/n$ (n is the number of SNPs used, $P < 4.06E-05$). To better compare trait-SNP GWAS results with the genes identified following DE analysis, the DE identified genes were transformed prior to analysis into a discontinuous pseudogenotype with two alleles, one if expression of the gene for a given line was higher than the median of all lines, and the other if it was lower. Three RIL populations were used to map QTL: BY population: BK (an inbred selected from a tropical landrace with very big kernel size) × Yu87-1 (an elite inbred with tropical background used frequently in Chinese breeding programs), SZ population: SK (an inbred selected from a tropical landrace with very small kernel size) × Zheng58 (an elite inbred used frequently in Chinese breeding programs), and YZ population: Yu87-1 (an elite inbred with tropical background used frequently in Chinese breeding programs) × Zong3 (an elite inbred used frequently in Chinese breeding programs). The QTLs were mapped for three flowering times traits: DTT, measured as the number of days from planting to 50% male flower appearance; DTS or the number of days from planting to 50% female flower appearance; and days to anthesis (DTA), the number of days from planting to 50% male flower pollen shed. Mapping was done using WinQTLcart (Wang et al., 2011) and a limit of detection threshold value for significance was set to 2.5. The flowering time traits are an indication of adaptation and thus used as a proxy for that trait.

Comparison with DE Gene Sets from Other Tissues

Li et al. (2012) conducted RNA-seq on the shoot apex from 2-week-old seedlings of the NAM founders. We downloaded the raw data and quantified with the same procedure used in the present study (Fu et al., 2013), and average expression level of the two runs of each line was used. The third leaf (L3) after germination of the NAM founders was also RNA-sequenced by the Springer laboratory, and the RPKM results were obtained by the author (Eichten et al., 2013). These two studies provided different tissues and different genetic backgrounds compared with the current study; these were used to test if genes identified with DE in the current study are tissue specific or if, as assumed in the current study, will prove to have lasting effects during maize development and maturity. Furthermore, consistency of expression differences in different populations can be validated. Genes with zero mapped reads in more than 60% of the inbred lines were excluded, and the overall distribution of expression levels for each remaining gene was normalized using the normal quantile transformation (Fu et al., 2013). Lines with unmixed temperate (NSS + SS) and tropical backgrounds were used to call DE genes ($P < 0.05$) with the same method (Fu et al., 2013). To test if the number of overlapping DE genes (Supplemental Figure 3) is significant between the current study and the two published experiments, random subsamples of genes were chosen from each comparison pair (using the same number for each subset as the number of DE genes identified by each study); this simulation was repeated 10 000 times to create a distribution of random overlap DE comparisons (Supplemental Figure 3). This distribution was used to test if the observed number of DE overlaps is in accordance with the simulated normal distribution.

REFERENCES

Ahmad, A., Zhang, Y., and Cao, X.-F. (2010). Decoding the epigenetic language of plant development. Mol. Plant **3**:719–728.

Antão, C.M., and Malcata, F.X. (2005). Plant serine proteases: biochemical, physiological and molecular features. Plant Physiol. Biochem. **43**:637–650.

Baldi, P., and Long, A.D. (2001). A Bayesian framework for the analysis of microarray expression data: regularized t-test and statistical inferences of gene changes. Bioinformatics **17**:509–519.

Beatty, M.K., Rahman, A., Cao, H., Woodman, W., Lee, M., Myers, A.M., and James, M.G. (1999). Purification and molecular genetic characterization of ZPU1, a pullulanase-type starch-debranching enzyme from maize. Plant Physiol. **119**:255–266.

Bhargava, S., and Sawant, K. (2013). Drought stress adaptation: metabolic adjustment and regulation of gene expression. Plant Breed. **132**:21–32.

Blake, M. (2006). Dating the Initial Spread of *Zea mays*. Histories of Maize: Multidisciplinary Approaches to the Prehistory, Linguistics, Biogeography, Domestication and Evolution of Maize (Amsterdam: Elsevier), pp. 55–72.

Bohnert, H.J., and Sheveleva, E. (1998). Plant stress adaptations–making metabolism move. Curr. Opin. Plant Biol. **1**:267–274.

Buckler, E.S., Holland, J.B., Bradbury, P.J., Acharya, C.B., Brown, P.J., Browne, C., Ersoz, E., Flint-Garcia, S., Garcia, A., Glaubitz, J.C., et al. (2009). The genetic architecture of maize flowering time. Science **325**:714–718.

Carroll, S.B. (2008). Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. Cell **134**:25–36.

Chander, S., Guo, Y., Yang, X., Zhang, J., Lu, X., Yan, J., Song, T., Rocheford, T.R., and Li, J. (2008). Using molecular markers to identify two major loci controlling carotenoid contents in maize grain. Theor. Appl. Genet. **116**:223–233.

Chen, Z.J. (2007). Genetic and epigenetic mechanisms for gene expression and phenotypic variation in plant polyploids. Annu. Rev. Plant Biol. **58**:377.

Chen, H., Patterson, N., and Reich, D. (2010). Population differentiation as a test for selective sweeps. Genome Res. **20**:393–402.

Chia, J.M., Song, C., Bradbury, P.J., Costich, D., de Leon, N., Doebley, J., Elshire, R.J., Gaut, B., Geller, L., and Glaubitz, J.C. (2012). Maize HapMap2 identifies extant variation from a genome in flux. Nat. Genet. **44**:803–807.

Ching, A., Dhugga, K.S., Appenzeller, L., Meeley, R., Bourett, T.M., Howard, R.J., and Rafalski, A. (2006). Brittle stalk 2 encodes a putative glycosylphosphatidylinositol-anchored protein that affects mechanical strength of maize tissues by altering the composition and structure of secondary cell walls. Planta **224**:1174–1184.

Cho, Y., Fernandes, J., Kim, S.H., and Walbot, V. (2002). Gene-expression profile comparisons distinguish seven organs of maize. Genome Biol. **3**, research0045.

Cockram, J., Thiel, T., Steuernagel, B., Stein, N., Taudien, S., Bailey, P.C., and O'Sullivan, D.M. (2012). Genome dynamics explain the evolution of flowering time CCT domain gene families in the Poaceae. PLoS One **7**:e45307.

Cortessis, V.K., Thomas, D.C., Levine, A.J., Breton, C.V., Mack, T.M., Siegmund, K.D., Haile, R.W., and Laird, P.W. (2012). Environmental epigenetics: prospects for studying epigenetic mediation of exposure–response relationships. Hum. Genet. **131**:1565–1589.

de Casas, R.R., Kovach, K., Dittmar, E., Barua, D., Barco, B., and Donohue, K. (2012). Seed after-ripening and dormancy determine adult life history independently of germination timing. New Phytol. **194**:868–879.

Dharmasiri, N., Dharmasiri, S., and Estelle, M. (2005). The F-box protein TIR1 is an auxin receptor. Nature **435**:441–445.

Dharmasiri, S., Jayaweera, T., and Dharmasiri, N. (2013). Plant hormone signalling: current perspectives on perception and mechanisms of action. Ceylon J. Sci. Biol. Sci. **42**:1–17.

Doebley, J.F., Gaut, B.S., and Smith, B.D. (2006). The molecular genetics of crop domestication. Cell **127**:1309–1321.

Du, Z., Zhou, X., Ling, Y., Zhang, Z., and Su, Z. (2010). agriGO: a GO analysis toolkit for the agricultural community. Nucleic Acids Res. **38**:64–70.

Dubreuil-Maurizi, C., and Poinssot, B. (2012). Role of glutathione in plant signaling under biotic stress. Plant Signal. Behav. **7**:210–212.

Ducrocq, S., Madur, D., Veyrieras, J.-B., Camus-Kulandaivelu, L., Kloiber-Maitz, M., Presterl, T., Ouzunova, M., Manicacci, D., and Charcosset, A. (2008). Key impact of Vgt1 on flowering time adaptation in maize: evidence from association mapping and ecogeographical information. Genetics **178**:2433–2437.

Eastmond, P.J. (2006). SUGAR-DEPENDENT1 encodes a patatin domain triacylglycerol lipase that initiates storage oil breakdown in germinating *Arabidopsis* seeds. Plant Cell **18**:665–675.

Ecker, J.R., Bickmore, W.A., Barroso, I., Pritchard, J.K., Gilad, Y., and Segal, E. (2012). Genomics: ENCODE explained. Nature **489**:52–55.

Eichten, S.R., Briskine, R., Song, J., Li, Q., Swanson-Wagner, R., Hermanson, P.J., Waters, A.J., Starr, E., West, P.T., and Tiffin, P. (2013). Epigenetic and genetic influences on DNA methylation variation in maize populations. Plant Cell **25**:2783–2797.

Fraser, H.B. (2013). Gene expression drives local adaptation in humans. Genome Res. **23**:1089–1096.

Fu, J., Cheng, Y., Linghu, J., Yang, X., Kang, L., Zhang, Z., Zhang, J., He, C., Du, X., Peng, Z., et al. (2013). RNA sequencing reveals the complex regulatory network in maize kernel. Nat. Commun. **4**:2832.

Ganal, M.W., Durstewitz, G., Polley, A., Bérard, A., Buckler, E.S., Charcosset, A., Clarke, J.D., Graner, E.M., Hansen, M., Joets, J., et al. (2011). A large maize (*Zea mays* L.) SNP genotyping array: development and germplasm genotyping, and genetic mapping to compare with the B73 reference genome. PLoS One **6**:e28334.

Griffiths, S., Dunford, R.P., Coupland, G., and Laurie, D.A. (2003). The evolution of CONSTANS-like gene families in barley, rice, and *Arabidopsis*. Plant Physiol. **131**:1855–1867.

Grobman, A., Bonavia, D., Dillehay, T.D., Piperno, D.R., Iriarte, J., and Holst, I. (2012). Preceramic maize from Paredones and Huaca Prieta, Peru. Proc. Natl. Acad. Sci. USA **109**:1755–1759.

Haas, J., Creamer, W., Mesía, L.H., Goldstein, D., Reinhard, K., and Rodríguez, C.V. (2013). Evidence for maize (*Zea mays*) in the Late Archaic (3000–1800 BC) in the Norte Chico region of Peru. Proc. Natl. Acad. Sci. USA **110**:4945–4949.

Harrell, F.E. (2012). Hmisc: Harrell Miscellaneous, Available online at: http://CRAN.R-project.org/package=Hmisc.

Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G., and Jarvis, A. (2005). Very high resolution interpolated climate surfaces for global land areas. Int. J. Climatol. **25**:1965–1978.

Hijmans, R.J., Guarino, L., and Mathur, P. (2012). DIVA-GIS Is a Free Computer Program for Mapping and Geographic Data Analysis, Available online at: http://www.diva-gis.org.

Hufford, M.B., Xu, X., Van Heerwaarden, J., Pyhäjärvi, T., Chia, J.M., Cartwright, R.A., Elshire, R.J., Glaubitz, J.C., Guill, K.E., and Kaeppler, S.M. (2012). Comparative population genomics of maize domestication and improvement. Nat. Genet. **44**:808–811.

Jain, M., Nijhawan, A., Arora, R., Agarwal, P., Ray, S., Sharma, P., Kapoor, S., Tyagi, A.K., and Khurana, J.P. (2007). F-box proteins in rice. genome-wide analysis, classification, temporal and spatial gene expression during panicle and seed development, and regulation by light and abiotic stress. Plant Physiol. **143**:1467–1483.

Jia, F., Wu, B., Li, H., Huang, J., and Zheng, C. (2013). Genome-wide identification and characterisation of F-box family in maize. Mol. Genet. Genomics **288**:559–577.

Jiao, Y., Zhao, H., Ren, L., Song, W., Zeng, B., Guo, J., Wang, B., Liu, Z., Chen, J., Li, W., et al. (2012). Genome-wide genetic changes during modern breeding of maize. Nat. Genet. **44**:812–815.

Kapazoglou, A., Drosou, V., Argiriou, A., and Tsaftaris, A.S. (2013). The study of a barley epigenetic regulator, HvDME, in seed development and under drought. BMC Plant Biol. **13**:172.

Karlsson, K., Sharma, H., and Nyberg, F. (2006). Chromatographic characterization of substance P endopeptidase in the rat brain reveals affected enzyme activity following heat stress. Biomed. Chromatogr. **20**:77–82.

Kepinski, S., and Leyser, O. (2005). The *Arabidopsis* F-box protein TIR1 is an auxin receptor. Nature **435**:446–451.

Khanna, R., Kikis, E.A., and Quail, P.H. (2003). EARLY FLOWERING 4 functions in phytochrome B-regulated seedling de-etiolation. Plant Physiol. **133**:1530–1538.

Kinsella, R.J., Kähäri, A., Haider, S., Zamora, J., Proctor, G., Spudich, G., Almeida-King, J., Staines, D., Derwent, P., Kerhornou, A., et al. (2011). Ensembl BioMarts: a hub for data retrieval across taxonomic space. Database (Oxford) **2011**, bar030.

Koenig, D., Jiménez-Gómez, J.M., Kimura, S., Fulop, D., Chitwood, D.H., Headland, L.R., Kumar, R., Covington, M.F., Devisetty, U.K., and Tat, A.V. (2013). Comparative transcriptomics reveals patterns of selection in domesticated and wild tomato. Proc. Natl. Acad. Sci. USA **110**:E2655–E2662.

Kolmos, E., Schoof, H., Plümer, M., and Davis, S.J. (2008). Structural insights into the function of the core-circadian factor TIMING OF CAB2 EXPRESSION 1 (TOC1). J. Circadian Rhythms **6**:3.

Koops, P., Pelser, S., Ignatz, M., Klose, C., Marrocco-Selden, K., and Kretsch, T. (2011). EDL3 is an F-box protein involved in the regulation of abscisic acid signalling in *Arabidopsis thaliana*. J. Exp. Bot. **62**:5547–5560.

Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., Jones, S.J., and Marra, M.A. (2009). Circos: an information aesthetic for comparative genomics. Genome Res. **19**:1639–1645.

Lawrence, C.J., Harper, L.C., Schaeffer, M.L., Sen, T.Z., Seigfried, T.E., and Campbell, D.A. (2008). MaizeGDB: the maize model organism database for basic, translational, and applied research. Int. J. Plant Genomics **2008**:496957.

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

Li, X., Zhu, C., Yeh, C.T., Wu, W., Takacs, E.M., Petsch, K.A., Tian, F., Bai, G., Buckler, E.S., Muehlbauer, G.J., et al. (2012). Genic and nongenic contributions to natural variation of quantitative traits in maize. Genome Res. **22**:2436–2444.

Li, H., Peng, Z., Yang, X., Wang, W., Fu, J., Wang, J., Han, Y., Chai, Y., Guo, T., Yang, N., et al. (2013). Genome-wide association study dissects the genetic architecture of oil biosynthesis in maize kernels. Nat. Genet. **45**:43–50.

Lipka, A.E., Tian, F., Wang, Q., Peiffer, J., Li, M., Bradbury, P.J., Gore, M.A., Buckler, E.S., and Zhang, Z. (2012). GAPIT: genome association and prediction integrated tool. Bioinformatics **28**:2397–2399.

Matsuoka, Y., Vigouroux, Y., Goodman, M.M., Sanchez, 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.

McCarthy, F.M., Wang, N., Magee, G.B., Nanduri, B., Lawrence, M.L., Camon, E.B., Barrell, D.G., Hill, D.P., Dolan, M.E., Williams, W.P., et al. (2006). AgBase: a functional genomics resource for agriculture. BMC Genomics **7**:229.

McMullen, M.D., Frey, M., and Degenhardt, J. (2009). Genetics and biochemistry of insect resistance in maize. In Handbook of maize: Its biology, J.L. Bennetzen and S.C. Hake, eds. (New York, USA: Springer), pp. 271–289.

Münster, T., Deleu, W., Wingen, L., and Ouzunova, M. (2002). Maize MADS-box genes galore. Maydica **47**:287–301.

Nei, M. (1977). F-statistics and analysis of gene diversity in subdivided populations. Ann. Hum. Genet. **41**:225–233.

Patterson, N., Price, A.L., and Reich, D. (2006). Population structure and eigenanalysis. PLoS Genet. **2**:e190.

Pavlidis, P., Jensen, J.D., Stephan, W., and Stamatakis, A. (2012). A critical assessment of storytelling: gene ontology categories and the importance of validating genomic scans. Mol. Biol. Evol. **29**:3237–3248.

Phelps-Durr, T.L., Thomas, J., Vahab, P., and Timmermans, M.C. (2005). Maize rough sheath2 and its *Arabidopsis* orthologue ASYMMETRIC LEAVES1 interact with HIRA, a predicted histone chaperone, to maintain knox gene silencing and determinacy during organogenesis. Plant Cell **17**:2886–2898.

Piperno, D.R., and Pearsall, D.M. (1998). Origins of Agriculture in the Lowland Neotropics (New York: Academic Press).

Price, A.L., Patterson, N.J., Plenge, R.M., Weinblatt, M.E., Shadick, N.A., and Reich, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. Nat. Genet. **38**:904–909.

Pritchard, J.K., Stephens, M., and Donnelly, P. (2000). Inference of population structure using multilocus genotype data. Genetics **155**:945–959.

Ptashne, M. (2007). On the use of the word 'epigenetic'. Curr. Biol. **17**:R233–R236.

Richen, C., Xiting, Z., and Junping, G. (2003). Effect of water deficit stress on endopeptidase activity in petals of cut rose 'Belami'. Acta Hort. Sin. **30**:352–354.

Rideout, W., Coetzee, G.A., Olumi, A.F., and Jones, P.A. (1990). 5-Methylcytosine as an endogenous mutagen in the human LDL receptor and p53 genes. Science **249**:1288–1290.

Robson, F., Costa, M.M., Hepworth, S.R., Vizir, I., Piñeiro, M., Reeves, P.H., Putterill, J., and Coupland, G. (2001). Functional importance of conserved domains in the flowering-time gene CONSTANS demonstrated by analysis of mutant alleles and transgenic plants. Plant J. **28**:619–631.

Schlebusch, C.M., Skoglund, P., Sjödin, P., Gattepaille, L.M., Hernandez, D., Jay, F., Li, S., De Jongh, M., Singleton, A., Blum, M.G., et al. (2012). Genomic variation in seven Khoe-San groups reveals adaptation and complex African history. Science **338**:374–379.

Schuster-Böckler, B., and Lehner, B. (2012). Chromatin organization is a major influence on regional mutation rates in human cancer cells. Nature **488**:504–507.

Sekhon, R.S., Lin, H., Childs, K.L., Hansey, C.N., Buell, C.R., de Leon, N., and Kaeppler, S.M. (2011). Genome-wide atlas of transcription during maize development. Plant J. **66**:553–563.

Sharma, M., Chai, C., Morohashi, K., Grotewold, E., Snook, M.E., and Chopra, S. (2012). Expression of flavonoid 3′-hydroxylase is controlled by P1, the regulator of 3-deoxyflavonoid biosynthesis in maize. BMC Plant Biol. **12**:196.

Shen, H., Zhu, L., Bu, Q.Y., and Huq, E. (2012). MAX2 affects multiple hormones to promote photomorphogenesis. Mol. Plant **5**:750–762.

Song, S., Dai, X., and Zhang, W. (2012). A rice F-box gene, OsFbx352, is involved in glucose-delayed seed germination in rice. J. Exp. Bot. **63**:5559–5568.

Staller, J.E., and Thompson, R.G. (2002). A multidisciplinary approach to understanding the initial introduction of maize into coastal Ecuador. J. Archaeol. Sci. **29**:33–50.

Stern, D.L., and Orgogozo, V. (2009). Is genetic evolution predictable? Science **323**:746–751.

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.

Tajima, F. (1983). Evolutionary relationship of DNA sequences in finite populations. Genetics **105**:437–460.

Tollenaar, M., and Wu, J. (1999). Yield improvement in temperate maize is attributable to greater stress tolerance. Crop Sci. **39**:1597–1604.

USDA FAS. (2013). Circular Series: World Cultural Production, Available online at: http://www.fas.usda.gov/psdonline/circulars/production.pdf.

Van Heerwaarden, J., Doebley, J., Briggs, W.H., Glaubitz, J.C., Goodman, M.M., Gonzalez, J.D.J.S., and Ross-Ibarra, J. (2011). Genetic signals of origin, spread, and introgression in a large sample of maize landraces. Proc. Natl. Acad. Sci. USA **108**:1088–1092.

Van Zanten, M., Koini, M.A., Geyer, R., Liu, Y., Brambilla, V., Bartels, D., Koornneef, M., Fransz, P., and Soppe, W.J.J. (2011). Seed maturation in *Arabidopsis thaliana* is characterized by nuclear size reduction and increased chromatin condensation. Proc. Natl. Acad. Sci. USA **108**:20219–20224.

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

Wang, Z., Gerstein, M., and Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. Nat. Rev. Genet. **10**:57–63.

Wang, S., Basten, C.J., and Zeng, Z. (2011). Windows QTL Cartographer 2.5 (Raleigh, NC: Department of Statistics, North Carolina State University), Available online at: http://statgen.ncsu.edu/qtlcart/WQTLCart.htm.

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

Wickham, H. (2009). ggplot2: Elegant Graphics for Data Analysis (New York: Springer)978-0-387-98140-6, Available online at: http://had.co.nz/ggplot2/book.

Wimmer, V., Albrecht, T., Auinger, H.J., and Schön, C.C. (2012). synbreed: a framework for the analysis of genomic prediction data using R. Bioinformatics **28**:2086–2087.

Wollmann, H., and Berger, F. (2012). Epigenetic reprogramming during plant reproduction and seed development. Curr. Opin. Plant Biol. **15**:63–69.

Wolters, H., and Jürgens, G. (2009). Survival of the flexible: hormonal growth control and adaptation in plant development. Nat. Rev. Genet. **10**:305–317.

Wray, G.A. (2007). The evolutionary significance of cis-regulatory mutations. Nat. Rev. Genet. **8**:206–216.

Xi, W., Liu, C., Hou, X., and Yu, H. (2010). MOTHER OF FT AND TFL1 regulates seed germination through a negative feedback loop modulating ABA signaling in *Arabidopsis*. Plant Cell **22**:1733–1748.

Xu, Z., and Dooner, H.K. (2006). The maize aberrant pollen transmission 1 gene is a SABRE/KIP homolog required for pollen tube growth. Genetics **172**:1251–1261.

Yan, J., Shah, T., Warburton, M.L., Buckler, E.S., McMullen, M.D., and Crouch, J. (2009). Genetic characterization and linkage disequilibrium estimation of a global maize collection using SNP markers. PLoS One **4**:e8451.

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

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

Yoo, S.Y., Kardailsky, I., Lee, J.S., Weigel, D., and Ahn, J.H. (2004). Acceleration of flowering by overexpression of MFT (MOTHER OF FT AND TFL1). Mol. Cells **17**:95–101.

Zdobnov, E.M., and Apweiler, R. (2001). InterProScan–an integration platform for the signature-recognition methods in InterPro. Bioinformatics **17**:847–848.

Zhang, H., and Ogas, J. (2009). An epigenetic perspective on developmental regulation of seed genes. Mol. Plant **2**:610–627.

Zhang, X., Colleoni, C., Ratushna, V., Sirghie-Colleoni, M., James, M., and Myers, A. (2004). Molecular characterization demonstrates that the *Zea mays* gene sugary2 codes for the starch synthase isoform SSIIa. Plant Mol. Biol. **54**:865–879.

Zhao, Q., Weber, A.L., McMullen, M.D., Guill, K., and Doebley, J. (2011). MADS-box genes of maize: frequent targets of selection during domestication. Genet. Res. **93**:65–75.