the plant journal

SEB Society for Experimental Biology

The Plant Journal (2016) **86,** 391–402 doi: 10.1111/tpj.13174

Identification of genetic variants associated with maize flowering time using an extremely large multi-genetic background population

Yong-xiang Li[1], Chunhui Li[1], Peter J. Bradbury[2], Xiaolei Liu[2], Fei Lu[2], Cinta M. Romay[2], Jeffrey C. Glaubitz[2], Xun Wu[1], Bo Peng[1], Yunsu Shi[1], Yanchun Song[1], Dengfeng Zhang[1], Edward S. Buckler[2,3,*], Zhiwu Zhang[4,5,*], Yu Li[1,*] and Tianyu Wang[1,*]

[1]*Institute of Crop Science, Chinese Academy of Agricultural Sciences, Beijing 10008, China,*
[2]*Institute for Genomic Diversity, Cornell University, Ithaca, NY 14853, USA,*
[3]*United States Department of Agriculture-Agricultural Research Service, Ithaca, NY 14853, USA,*
[4]*Department of Animal Science, Northeast Agricultural University, Harbin, Heilongjiang 150030, China, and*
[5]*Department of Crop and Soil Sciences, Washington State University, Pullman, WA 99164, USA*

Received 9 August 2015; revised 17 March 2016; accepted 18 March 2016; published online 25 March 2016.
*For correspondence (e-mails esb33@cornell.edu; zhiwu.zhang@wsu.edu; liyu03@caas.cn; wangtianyu@caas.cn).

SUMMARY

Flowering time is one of the major adaptive traits in domestication of maize and an important selection criterion in breeding. To detect more maize flowering time variants we evaluated flowering time traits using an extremely large multi- genetic background population that contained more than 8000 lines under multiple Sino-United States environments. The population included two nested association mapping (NAM) panels and a natural association panel. Nearly 1 million single-nucleotide polymorphisms (SNPs) were used in the analyses. Through the parallel linkage analysis of the two NAM panels, both common and unique flowering time regions were detected. Genome wide, a total of 90 flowering time regions were identified. One-third of these regions were connected to traits associated with the environmental sensitivity of maize flowering time. The genome-wide association study of the three panels identified nearly 1000 flowering time-associated SNPs, mainly distributed around 220 candidate genes (within a distance of 1 Mb). Interestingly, two types of regions were significantly enriched for these associated SNPs – one was the candidate gene regions and the other was the approximately 5 kb regions away from the candidate genes. Moreover, the associated SNPs exhibited high accuracy for predicting flowering time.

Keywords: maize (*Zea mays* L.), flowering time, genome-wide association study (GWAS), linkage analysis, nested association mapping (NAM).

INTRODUCTION

Since the domestication of maize from Balsas teosinte (*Zea mays* ssp. *parviglumis*) in the Mexican highlands approximately 9000 years ago, the area planted to this crop has been expanding (Matsuoka *et al.*, 2002). During its spread throughout the world, maize has evolved to adapt to diverse ecological conditions. Flowering time reflects a plant's adaptive response to its environment through floral transition to local conditions and is considered a major adaptive trait and an important selection criterion in plant breeding (Jung and Müller, 2009). Thus, post-domestication and breeding selection have driven the plant's diverse range of flowering times, which optimize growth under given local conditions (Elzinga *et al.*, 2007).

Flowering time has been extensively studied in several plant species (Izawa *et al.*, 2003; Andersen *et al.*, 2004; Jung and Müller, 2009; Kong *et al.*, 2010). Hundreds of flowering time genes have been identified in Arabidopsis and used to construct the genetic regulatory network (GRN) (Ehrenreich *et al.*, 2009; Jung and Müller, 2009; Brachi *et al.*, 2010; Chen *et al.*, 2012). Maize exhibits high levels of genetic diversity (Wright *et al.*, 2005) and a wide variety of flowering times, ranging from 35 to 120 days (Colasanti and Muszynski, 2009). Recently, several major genetic components that regulate maize flowering time, for example *Vgt1* (Salvi *et al.*, 2007) and *ZmCCT* (Hung *et al.*, 2012; Yang *et al.*, 2013), have been positionally

cloned. A conceptual GRN model for maize flowering time that includes more than 40 genes was also proposed using data from maize, grass and Arabidopsis (Dong *et al.*, 2012). However, many more genes are involved, as demonstrated by the 62 consensus quantitative trait loci (QTLs) that have been identified in maize from multiple genetic backgrounds (Chardon *et al.*, 2004). Therefore, additional genetic components must be dissected and fitted to produce a more complete and accurate GRN model of maize flowering time.

Previously, a large nested association mapping (NAM) population, including 5000 recombinant inbred lines (RILs) from 25 crosses between diverse inbred lines and a common parent (B73), was used to investigate the genetic architecture of maize flowering time (Buckler *et al.*, 2009). The results of this study provided evidence that maize flowering time can be predicted by a model of numerous small additive QTLs. Thus, a diverse germplasm panel that include as many variants of these genes as possible is crucial for the thorough dissection of the genetic architecture of flowering time and its prediction in breeding programs.

In this study, we used three maize germplasm panels: (i) the highly genetically diverse NAM population with 5000 RILs described above (US-NAM) developed in the United States, (ii) another genetically independent maize NAM population with 2000 RILs developed in China (CN-NAM), and (iii) a natural association panel with 1745 inbred lines (Ames) to identify the genetic components of maize flowering time. Our ultimate objective was to identify new genetic information to support construction of an improved flowering time GRN, thereby increasing the prediction accuracy of this complex trait.

RESULTS

Phenotyping of flowering time

Maize flowering time was measured for both male and female flowers as days to anthesis (DA) and days to silking (DS), respectively. The anthesis–silking interval (ASI) was derived from the difference between DA and DS. These flowering time traits were measured for US-NAM, CN-NAM and Ames in multiple environments, defined as combinations of years and locations.

The US-NAM consists of 5000 RILs from 25 original crosses between diverse inbred lines and a common parent, B73, and was developed in the United States. Among the 13 environments evaluated for the US-NAM, eight were measured in the United States and reported previously (Buckler *et al.*, 2009). The other five environments were measured in China. The environmental locations varied from subtropical to temperate conditions (Data S1 in Supporting Information).

The flowering time measurements under multiple environments were analyzed in two ways. First, we used the best linear unbiased predictions (BLUPs) of the US-NAM RILs, derived across 13 environments, with a mixed model (Brown *et al.*, 2011). These BLUPs were used as the overall performance of each RIL across environments. The broad sense heritabilities of DA, DS and ASI of the US-NAM across all 13 environments were 93, 92 and 80%, respectively. Second, we used the sensitivities of RILs across environments, which were defined as variation in flowering time among the 13 environments, to reflect the environmental response of each US-NAM RIL. We derived the coefficient of variation (CV) for both DA (DACV) and DS (DSCV). These coefficients varied from 12.41 to 17.15% for DACV and from 13.17 to 17.40% for DSCV (Data S2). High correlation was observed between DACV and DSCV (R^2 = 0.61). We observed that RIL families with tropical parents had higher variation in flowering time than those from temperate parents. The DACV and DSCV means for tropical RILs were 16.1 and 16.5%, respectively. Comparatively, the DACV and DSCV means for temperate RILs were 14.8 and 14.3%, respectively.

The CN-NAM consists of 2000 RILS from the crosses between 11 diverse inbred lines and a common parent (HZS) (Li *et al.*, 2013). All 12 parents of the CN-NAM originated from a minimum core collection of maize inbred lines in China (Wang *et al.*, 2008). Compared with the US-NAM, the common parent and the majority of the other parents of the CN-NAM were from local Chinese temperate germplasm. The number of RILs in CN-NAM families ranged from 151 to 184, with an average of 178 (Data S3). These RILs were phenotyped at three locations across 2 years, for a total of six environments (Data S1). Similar to the US-NAM, ASI was derived from DA and DS. BLUPs and CVs of RILs were estimated across the six environments. Broad sense heritabilities were 91%, 90% and 86% for DA, DS, and ASI, respectively. The family means of CVs ranged from 14.35 to 17.03% for DACV and from 15.31 to 18.99% for DSCV (Data S3). We observed a strong correlation between DACV and DSCV (R^2 = 0.71).

The Ames panel consists 1745 inbred lines from the North Central Regional Plant Introduction Station (NCRPIS) of the US Department of Agriculture's Agricultural Research Service (USDA-ARS), located in Ames, IW, USA. The BLUPs on growing degree days DS (GDD-DS) of the Ames lines were previously estimated across three environments by Romay *et al.* (2013). Flowering times for these Ames lines were measured in Beijing, China, in 2012 (Data S1). We observed a strong correlation between DS measured in China and GDD-DS measured in the US (R^2 = 0.81).

Genotyping of germplasm

The genotypes of CN-NAM (http://www.panzea.org/ or http://www.cgris.net/maize/data1/index.htm), US-NAM and Ames (http://www.panzea.org/) contained a total of 0.95

million single-nucleotide polymorphisms (SNPs). These genotypes were obtained from the platform of genotyping-by-sequencing (GBS) in GBS BUILD v.2.7 by the Buckler Lab at Cornell University (http://www.maizegenetics.net/). Using the GBS data, Li *et al.* (2015) constructed high-density genetic maps, which consisted of 4932 bin markers for CN-NAM and 5296 bin markers for US-NAM. An evaluation of relationships among the founder lines of the two NAM panels revealed a distinct genetic architecture for each common parent and the other founder inbred lines (Rodgers-Melnick *et al.*, 2015). Thus, the CN-NAM represents the genetic diversity of the Chinese local germplasm and possesses genetic characteristics that are distinct from those of the US-NAM (Wang *et al.*, 2008; Li *et al.*, 2013; Rodgers-Melnick *et al.*, 2015).

Parallel linkage analyses identified common and specific QTLs

The QTLs of all five flowering time traits (DA, DS, ASI, DACV, DSCV) for each RIL family were initially mapped using the inclusive composite interval mapping (ICIM) method (Wang, 2012). Using ICIM, a total of 173 QTLs for CN-NAM and 480 QTLs for US-NAM were identified (Figure S1, Data S4).

We conducted QTL mapping (Figure S2) by using the joint stepwise regression (JSR) method (Buckler *et al.*, 2009). The identified QTLs were further evaluated by a two-dimensional examination approach (Broman and Sen, 2009). This approach tested whether a short chromosome region (<10 Mb) with a high logarithm of odds (LOD) value in JSR should be treated as one functional unit or as neighboring multi-units. Through this process, we were able to distinguish two major QTLs on chromosome 8 within a region of less than 9 Mb for both CN-NAM and US-NAM (Figure S3). Previously, this region was hypothesized to contain two linked QTLs (Buckler *et al.*, 2009).

For US-NAM, we identified 48, 46 and 39 QTLs that explained 89%, 88% and 63% of the total phenotypic variance for DA, DS and ASI, respectively, by using a full additive QTL model (Table 1). Compared with the mapping results of low-density markers (Buckler *et al.*, 2009), our results explained a similar phenotypic variance, but identified about 30% more QTLs. For CN-NAM, the total number of identified QTLs for DA, DS and ASI was approximately 60% of the total number of QTLs identified in US-NAM (Table 1). The overlaps of confidence intervals were 62, 68 and 37% between CN-NAM and US-NAM for DA, DS and ASI, respectively. The analysis of traits associated with environmental sensitivity in CN-NAM revealed 10 QTLs each for DACV and DSCV that explained 44 and 51% of the total phenotypic variation, respectively. There were 15 DACV and 13 DSCV QTLs identified in the US-NAM. These QTLs explained 46% and 45% of the total variation for DACV and DSCV, respectively. Four of the QTLs overlapped between CN-NAM and US-NAM.

Table 1 Number of quantitative trait loci (QTL) identified in CN-NAM and US-NAM

Trait	CN-NAM	US-NAM	Overlap
DA	29	48	18
DS	25	46	17
ASI	22	39	8
DACV	10	15	4
DSCV	10	13	4
Flowering time region	55	75	40

Three traits were derived from days to anthesis (DA) and days to silking (DS). The anthesis–silking interval (ASI) was defined as DA − DS. The coefficients of variation of DA and DS are denoted as DACV and DSCV, respectively.

According to the overlapped QTLs across five traits, 55 flowering time-related regions were identified for CN-NAM and 75 for US-NAM (Data S5). There were 40 regions overlapped between the two NAM panels. The rest of the regions were CN-NAM specific (15) and US-NAM-specific (35). Additionally, 29 regions, including the region containing the well-known photoperiod-related gene *ZmCCT* on chromosome 10, were identified to be associated with the environmental sensitivity of maize flowering time.

It is a common method to measure QTL pleiotropy as the overlap of QTLs among different traits. We investigated the overlapped QTLs among different traits in CN-NAM and US-NAM separately (Figure 1, Data S5). We found that more than 83 and 89% of the flowering QTLs were shared by at least two flowering time traits in CN-NAM and US-NAM, respectively. Six CN-NAM and 25 US-NAM QTLs were shared by at least three flowering traits. There were five QTL regions (one for CN-NAM and four for US-NAM) shared by all five traits. We also found that most of the environmental response QTLs were shared with flowering time QTLs. There were only a few QTLs that are specific for DACV (four for CN-NAM and zero for US-NAM) and DSCV (four for CN-NAM and one for US-NAM). This result suggests a dependent regulatory mechanism for the environmental response of maize flowering time.

Genome-wide association studies of CN-NAM, US-NAM and Ames for flowering time traits

Genome-wide association study (GWAS) analyses were performed to test the marker–trait associations of each flowering time trait for each panel. The analyses were repeated with partial random sampling using a bootstrap strategy. By using a bootstrap posterior probability (BPP) criterion of 0.05 (Valdar *et al.*, 2006; Tian *et al.*, 2011), we detected a total of 2501 marker–trait associations (CN-NAM 851, US-NAM 999, Ames 651) (Data S6). Under a stricter threshold of BPP $\geq$ 0.20 for single SNPs or a 5-Mb region with more than one SNP with BPP < 0.2 but $\geq$0.05, we identified 950 flowering time-associated SNPs (CN-NAM 340,

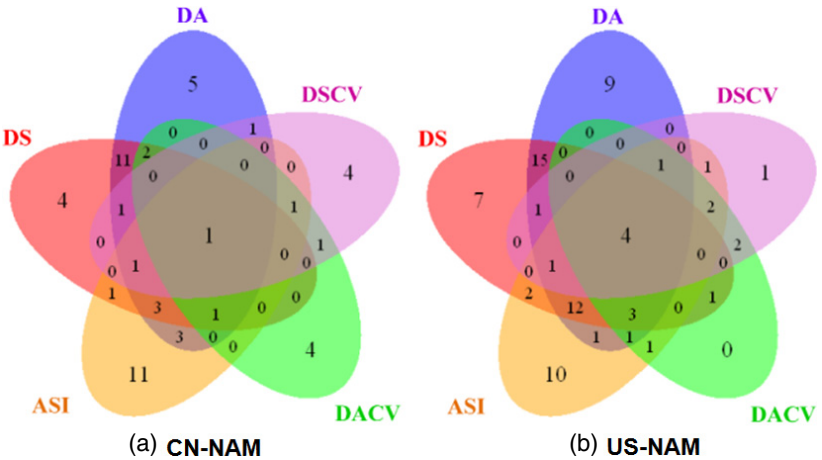

Figure 1. Number of quantitative trait loci (QTLs) shared among five flowering time traits.The QTL analyses were performed in CN-NAM (a) and US-NAM (b) separately. The five flowering time traits are (1) days to anthesis (DA), (2) days to silking (DS), (3) anthesis–silking interval (ASI), (4) coefficient of variation of DA (DACV) and (5) coefficient of variation of DS (DSCV).

US-NAM 367, Ames 243) (Figure 2). Although the associated SNPs clustered within very short regions and barely overlapped across different populations (only four SNPs were shared by CN-NAM and US-NAM; only two SNPs were shared by US-NAM and Ames), they confirmed the results of linkage analyses well.

Most of the associated SNPs were overlapped with the QTLs with $P < 0.001$ (Figure 3). For CN-NAM, US-NAM and Ames, 213 (63%), 261 (71%) and 128 (53%) of the associated SNPs, respectively, were distributed within the CN-NAM and US-NAM flowering time QTL regions. Among all CN-NAM and US-NAM QTLs, only two did not include associated SNPs. There were 31 QTLs (15 in CN-NAM and 16 in US-NAM, with 11 overlapping) that contained at least 10 associated SNPs. Therefore, we obtained consistent results from QTL mapping and GWAS.

Cross-validation

To evaluate how well a phenotype in one panel can be predicted by the associated SNPs identified in another panel, we conducted cross-validation among the three panels. For example, to evaluate the associated SNPs of CN-NAM in US-NAM, we re-estimated the effect of these SNPs in US-NAM with five-fold cross-validation. For each cross-validation test, 80% of US-NAM RILs were treated as the reference group and the other 20% were treated as the inference group. The effects of SNPs were estimated in the reference group by using the random-regression best linear unbiased prediction (RR-BLUP) method (Endelman, 2013). The estimated effects were used to predict the phenotypes of RILs in the inference group to evaluate the accuracy of prediction as the correlation between observed and predicted values. We repeat the process to another 20% of RILs as inference group, until all individual had their predictions. We repeated the process 100 times. For the purpose of evaluating model fit, the validation of a panel by its own associated SNPs was also performed using the same process. The results are illustrated in Figure S4.

We observed that the associated SNPs of US-NAM (367 SNPs) and Ames (243 SNPs) explained each other well (mean R^2 of 59% for US-NAM by Ames and 51% for Ames by US-NAM). This result is expected because the NAM founders are part of the Ames panel. In contrast, we observed that the associated SNPs of the US-NAM and Ames panels poorly explained CN-NAM (mean R^2 of 19% by US-NAM and 21% by Ames), and vice versa. When the associated SNPs of CN-NAM (340 SNPs) were fitted to US-NAM and Ames we also observed relatively lower genetic contributions (mean R^2 of 51% for US-NAM and 39% for Ames).

Independent validation

We used a published dataset (Romay *et al.*, 2013) that had 1515 lines overlapped with our data. The published data contained genotypes and GDD-DS that is highly correlated to DS. There also were additional 835 inbred lines, not contained in the Ames panel (Romay *et al.*, 2013), that can be used to perform an independent validation. Among the 835 lines, 401 lines were mainly composed of tropical lines. The other 434 lines appeared to exhibit a population structure similar to that of the 1515 lines in the Ames panel (Figure S5).

The independent validation was performed in the similar way to the cross-validation, during which the overlapped 1515 Ames lines were treated as the reference group and the independent 835 lines were treated as the inference group. We predicted the phenotypes of 835 lines by using the 243 associated SNPs identified from the 1515 Ames lines. We then compared the predictions with the actual observations. We observed an R^2 of 0.54 in a linear correlation between the observations and the predictions across the 835 lines. As expected, we found that the 243 associated SNPs of Ames could predict the 434 structurally related lines reasonably well ($R^2 = 51\%$). But, these SNPs were unable to predict the 401 tropical lines ($R^2 = 5\%$) which are distantly related to the Ames panel (Figure S6).

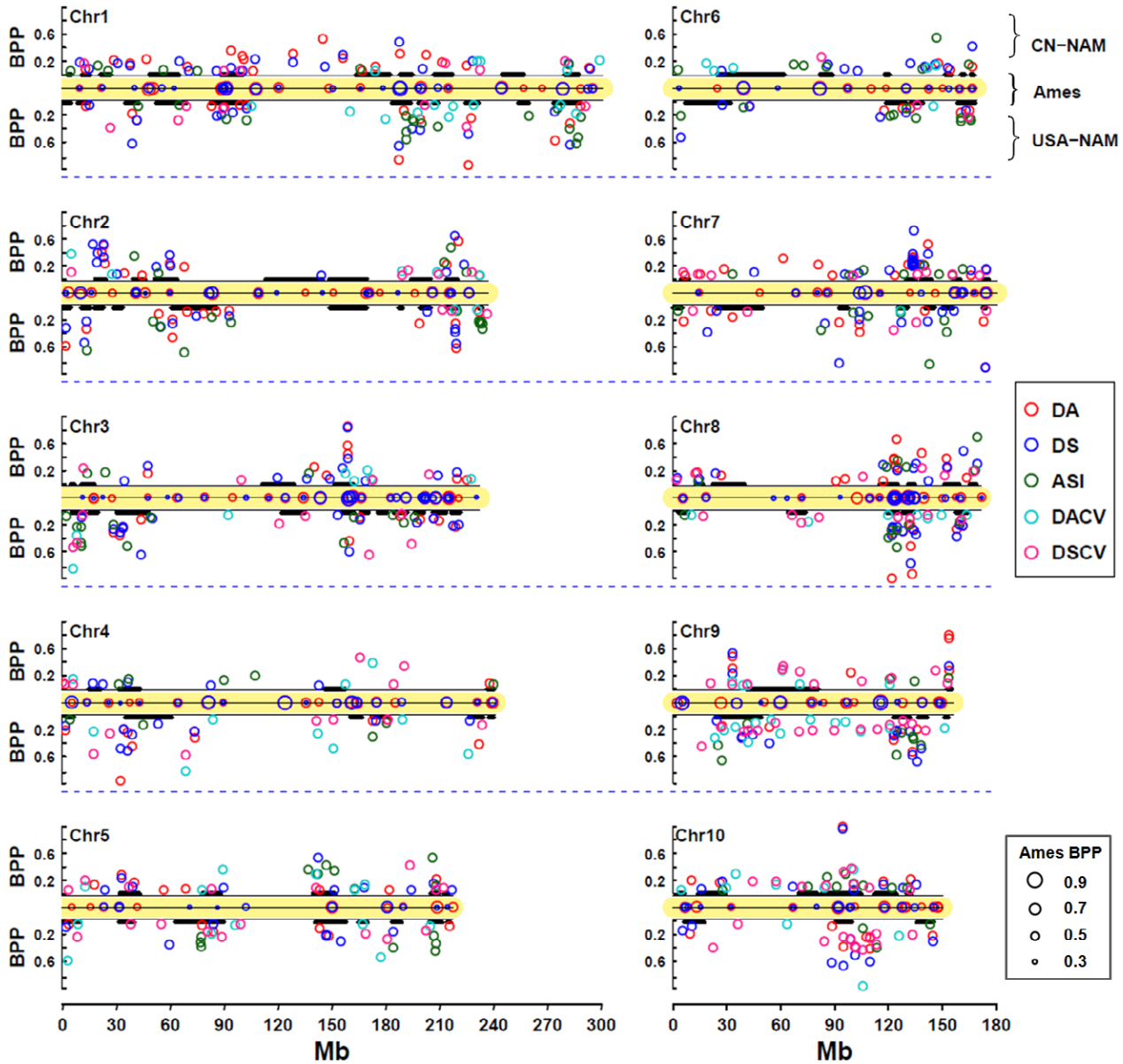

Figure 2. Overlap between quantitative trait loci (QTLs) and the associated single nucleotide polymorphisms.The linkage analyses were performed on CN-NAM and US-NAM populations. Association studies were performed on the Ames panel and the two NAM panels. The strength of association is demonstrated as the bootstrap posterior probability (BPP). The locations of QTLs and associated loci are displayed for each chromosome, trait and panel. BPP is illustrated by the scale on the upper panel for CN-NAM and the lower panel for US-NAM, and circle size in the middle panel for Ames. BPP is differentiated by color for different traits. Flowering time-related regions of all traits are presented by black solid lines in the upper and lower panels for CN-NAM and US-NAM, respectively.

Enrichment in maize flowering time candidate genes

Based on a literature search, a total of 919 maize flowering time candidate genes (Data S7), including cloned genes and homologs of other plants, were preliminarily selected to conduct enrichment analyses (Danilevskaya *et al.*, 2008; Chen *et al.*, 2012; Dong *et al.*, 2012; Hung *et al.*, 2012). Annotations of all selected genes were transferred from MaizeGDB (http://www.maizegdb.org/). We calculated the distances between the associated SNP and the candidate genes. The null distribution was derived from the distance between these associated SNPs and the same number of genes selected randomly from each chromosome. One thousand replicates were performed to derive the null distribution.

Notably, there were two regions that were enriched the most. One is the genic regions of the flowering time candidate genes. The other is the region about 5 kb away from the genic regions of candidates (Figure 4). We found that a total of 20 associated SNPs fell into the genic regions of candidates, including *RAP2.7*, which encodes an AP2-like

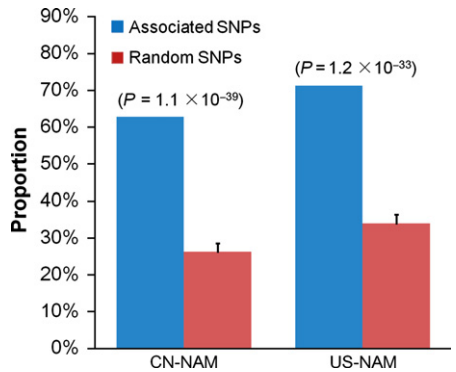

Figure 3. Enrichments of associated single nucleotide polymorphisms (SNPs) on flowering time quantitative trait loci (QTLs).The enrichments were conducted in CN-NAM and US-NAM separately. The enrichments are demonstrated as the difference between observed (blue) and expected (red) proportions of associated loci landing within QTLs for all traits. The expected proportion was derived from randomly selected markers in an amount equal to that of the observed loci. The random process was replicated for 1000 times. The standard errors were used for statistical tests and are displayed for the corresponding expectations.

transcription factor (Salvi *et al.*, 2007). We also found that 21 associated SNPs fell into the 5-kb non-genic region of candidates, including the well-known *ZCN8* (located 3.2 kb downstream, with a BPP of 0.97 in Ames) (Meng *et al.*, 2011).

The enriched causal candidates were expected to be responsible for the phenotypic polymorphism. In our study, a total of 220 candidates were hit within a 1-Mb region of associated SNPs (Data S8). Among them, 64% (140) were hit by at least two panels and 14% (31) were hit by all three panels. We also observed that 66% of the 220 candidates (145) were homologous with flowering time genes of Arabidopsis (Chen *et al.*, 2012). Additionally, from the 220 candidates, we identified 17 of the 45 genes within a proposed flowering time GRN. These 17 included genes involved in the autonomous, integrator, photoperiod, circadian clock and light transduction pathways (Dong *et al.*, 2012) (Data S8).

DISCUSSION

Flowering time variants largely determine environmental sensitivity in maize

A steady flowering time, which reflects insensitivity to environmental changes, is an important selection criterion in maize breeding. Generally, photoperiod greatly affects the environmental response of maize flowering time (Buckler *et al.*, 2009; Hung *et al.*, 2012; Yang *et al.*, 2013). Nevertheless, it is still difficult to distinguish the effects of other environmental factors on flowering time (Chen *et al.*, 2014). Therefore, our study calculated the CVs of DA and DS to estimate the general response of maize flowering time to photoperiod and other environmental factors (accumulated temperature, rainfall, etc.). Our objective was to provide additional clarity about the genetic architecture of this complex trait.

ZmCCT, a homolog of the rice photoperiod response gene *Ghd7* (Xue *et al.*, 2008), has been proposed as one of the most important genes affecting photoperiod response in maize (Hung *et al.*, 2012). A transposable element (TE) within the *ZmCCT* promoter can markedly reduce flowering time and photoperiod sensitivity (Yang *et al.*, 2013). Among the parents of US-NAM, our results revealed four (CML228, CML227, Ki11, and Ky21) with *ZmCCT* alleles (Hap4 and Hap5, without TE insertion) that increase flowering time and photoperiod sensitivity (Data S2) (Yang *et al.*, 2013). With the exception of the RIL family B73 × Ky21, the other three RIL families (B73 × CML228, B73 × CML277

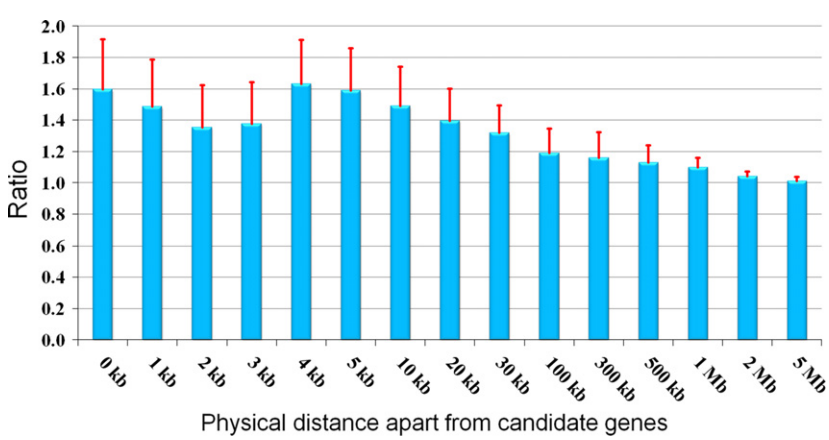

Figure 4. Enrichment of associated single nucleotide polymorphisms (SNPs) on flowering time candidate genes.In total, 950 associated SNPs were identified and 919 flowering time candidate genes were selected. Enrichment was defined as the ratio of observed to expected numbers of paired associated SNPs and nearby candidate genes within a specified distance. The expected numbers were derived from the same associated loci and randomly selected candidate genes in an amount equal to that of the true candidate genes. The random process was performed 1000 times. The numbers of paired associated loci and nearby candidate genes were recorded for distances varying from 0 to 5 Mb. Standard errors were calculated and are displayed correspondingly.

and B73 × Ki11) exhibited obvious environmental sensitivity and high flowering time variation, ranking first, third and sixth for DSCV, respectively, among all US-NAM RIL families. For these three families, we also obtained linear correlations between DS and DSCV ($R^2 > 0.1$) for their RILs, thereby indicating that *ZmCCT* can simultaneously affect both flowering time and environmental sensitivity. For all 12 parents of CN-NAM, only the inbred line of QI319 had tropical pedigrees (Teng *et al.*, 2004). This line possessed the sensitive *ZmCCT* allele and also exhibited the highest environmental sensitivity in terms of both DA and DS (Data S3).

In addition to those with sensitive *ZmCCT* alleles, other RIL families also exhibited both a long flowering time and high environmental sensitivity. For example, the RIL family B73 × CML52 had the longest flowering time, the highest DACV value and the second highest DSCV value. Consequently, more genes likely play a role in the performance of environmental sensitivity in maize and in the adaptive response of maize to diverse climatic effects. Moreover, we hypothesized that these genes either simultaneously influence flowering time and its environmental sensitivity (similar to *ZmCCT*) or only affect one of these traits.

Based on the integrated results of our two NAM panels, a total of 29 environmental sensitivity-related flowering time regions were detected (Data S5). However, we observed that most of these regions overlapped with the QTLs of the other three flowering time traits, DA, DS and ASI (Figure 1). Thus, we suggest that the genetic components of DA or DS play a determining role in the initiation of maize flowering time and also largely affect its environmental response.

Among the mapped QTL, we observed five pleiotropic loci that affected all five flowering time traits (Figure 1, Data S5). These loci included two major QTLs that cover the well-known maize flowering time genes *ZCN8* on chromosome 8 (Meng *et al.*, 2011) and *ZmCCT* on chromosome 10 (Hung *et al.*, 2012; Yang *et al.*, 2013). Moreover, we observed another QTL related to all five flowering time traits on chromosome 9 that appeared to exert marked effects on DACV and DSCV for both CN-NAM and US-NAM (Figure S2). We hypothesized that this QTL contained another major component or gene cluster that affects both maize flowering time and its environmental response.

At least 90 genetic components determine maize flowering time diversity

High heritabilities were obtained for the flowering time traits of both CN-NAM and US-NAM populations across multiple environments. Therefore, we expected the BLUPs across all identified environments to greatly decrease the effects of genotype × environment (g × e) interaction, allowing us to compare mapped results between CN-NAM and US-NAM. Applying a joint linkage mapping strategy to the NAM population, using the nested family and marker model, provides high dissection power for complex quantitative traits – particularly when marker density is high (Buckler *et al.*, 2009). In this study, 55 and 75 flowering time regions (or QTL clusters) for CN-NAM and US-NAM, respectively, were identified. Forty of these regions were found in both NAM panels, resulting in a total of 90 flowering time regions for the two panels combined. Therefore, 27% of the CN-NAM flowering time regions (15 of 55) and 47% of the US-NAM flowering time regions (35 of 75) were panel-specific. Consequently, we concluded that both common and specific flowering time regions determine maize flowering time for the two NAM panels.

Integration of the mapping results (both QTL mapping and GWAS) among the three independent panels provided both validation and complementary information for the dissection of the genetic architecture of maize flowering time. We observed multiple regions that contained both the QTLs and the intensively clustered associated SNPs across the three panels, including those regions with well-known flowering time components (*ZCN8* and *Vgt1* on chromosome 8 and *ZmCCT* on chromosome 10) (Meng *et al.*, 2011; Hung *et al.*, 2012; Yang *et al.*, 2013).

Marker-assisted recurrent selection (MARS) or genome selection (GS) is very helpful in crop breeding (Johnson, 2004). Thus, whether flowering time-associated SNPs can be used in the prediction of maize flowering time is an important question. In this study, the detected associated SNPs appeared to exhibit high predictive power for flowering time of genotypes with similar genetic backgrounds. In contrast, their predictive power for the tropical lines was poor, but may have been partly caused by a deviation in the breeding value estimation of the SNPs. This deviation could have originated from the limited freedom of the reference group or the existence of new rare variants in the tropical lines not contained by the Ames panel.

The enriched candidate genes provide valuable information for the construction of flowering time GRN in maize

Growing evidence supports the view that mutations in gene regulatory regions play significant roles in functional variants (Wray, 2007; Schaub *et al.*, 2012). In maize, Wallance *et al.* (2015) found that most variance could be explained by genic (within the gene region) and gene-proximal SNPs (at about 1–5 kb away from genes and likely positions of promoters and other short-range regulatory elements). In our study, we observed two obvious enrichment peaks within the genic and about 5-kb proximal regions of the candidate genes (Figure 4). Therefore, in addition to mutations within candidates, our findings provide additional evidence that variants at about 5-kb regions away from genes also significantly affect the performance of maize flowering time.

Hundreds of flowering time genes have been studied in plants. In Arabidopsis, the *FLOWERING LOCUS T* (*FT*) and *TERMINAL FLOWER1* (*TFL1*) genes, which constitute a type of *phosphatidylethanolamine-binding* (*PEBP*) gene, play important roles in floral transition (Fornara *et al.*, 2010; Pin and Nilsson, 2012). In our study, among the 24 *PEBP*-like genes in maize (Izawa *et al.*, 2003), nine were hit within the 1-Mb region of the flowering time-associated SNPs locations (the distance of the SNP from the candidate ranged from 3.2 to 670.2 kb and averaged 324.5 kb). These hits included six *FT*-like genes (*ZCN8*, *ZCN13*, *ZCN16*, *ZCN20*, *ZCN24*, *ZCN26*), two *TFL*-like genes (*ZCN5*, *ZCN6*) and one *MFT*-like gene (*ZCN10*).

The genes related to photoperiod play important roles in the expression of *FT*-like genes in a conceptual GRN model for maize flowering time (Dong *et al.*, 2012). The CONSTANS (CO) protein, which strongly influences the performance of maize flowering time in response to photoperiod, directly induces the transcription of *FT*-like genes in Arabidopsis (Srikanth and Schmid, 2011; Pin and Nilsson, 2012). Eight *CO*-like candidates were hit within a 1-Mb region of the locations of flowering time-associated SNPs (the distance of the SNP from candidates ranged from 0 to 912.7 kb and averaged 174.9 kb). These hits included *ZmCCT* (Hung *et al.*, 2012; Yang *et al.*, 2013), *CONZ1* (Miller *et al.*, 2008; Dong *et al.*, 2012; Hung *et al.*, 2012), *COL3*, *COL6* and *COL7* (Yilmaz *et al.*, 2009) and three homolog candidates (GRMZM2G176173, GRMZM2G 148772, GRMZM2G004483).

Gibberellin (GA) is an endogenous plant growth hormone that can affect *FT*-like gene expression under both long-day and short-day conditions (Osnato *et al.*, 2012; Song *et al.*, 2012). Considering the GA pathway, three MYB-like genes (*MYB74*, *MYB98* and *MYB69*) that positively regulate the expression of the GA biosynthesis gene, *GA20ox1*, in Arabidopsis (Song *et al.*, 2012) and five GA receptor-like candidates (GRMZM2G012546, GRMZM2G 173630, GRMZM2G406014, GRMZM2G006716, and GRMZM2G164454) were also hit within a 1-Mb region of flowering time-associated SNPs (the distance of the SNP from candidates ranged from 46.4 to 676.9 kb and averaged 243.1 kb).

MADS-box genes encode a family of transcription factors that take part in diverse developmental processes in flowering plants. We identified 10 MADS-box genes (*MADS9*, *MADS20*, *MADS21*, *MADS41*, *ZMM4*, *ZMM5*, *ZMM6*, *ZMM29*, *ZAG1* and GRMZM2G099522) (associated distance of the SNP from candidates ranged from 4.1 to 807.5 kb and averaged 339.9 kb). Among these genes *ZMM4*, a homolog of *SUPPRESSOR OF OVEREXPRESSION OF CONSTANS* (*SOC1*) in Arabidopsis (Malcomber *et al.*, 2006) and *OsMADS56* in rice (Ryu *et al.*, 2009), has been hypothesized to participate in the floral transition in the flowering time GRN in maize (Dong *et al.*, 2012). The *ZAG1* gene has been hypothesized to participate in floral development in maize (Thompson, 2009).

For the other flowering-regulating factors, we also hit seven *EARLY FLOWERING*-like candidates connected with maize flowering time (GRMZM2G401342, GRMZM2G 161913, GRMZM2G008765, GRMZM2G045275, AC233870. 1_FG003, GRMZM2G025646, GRMZM2G359322) (the associated SNP distance ranged from 0 to 686.8 kb and averaged 226.7 kb). The translation products of these candidate genes are directly bonded to the GI protein (Kim *et al.*, 2013). Three clock pathway genes (GRMZM2G033962, *ZmPRR37* and *ZmPRR73*), two light transduction pathway genes (*PHYB1* and *PHYB2*) (Dong *et al.*, 2012) and eight *SQUAMOSA PROMOTER BINDING PROTEIN*-like (*SPL*) (Wu *et al.*, 2009) aging pathway genes (*LG1*, *SBP3*, *SBP8*, *SBP17*, *SBP20*, *SBP27*, *SBP29* and *TSH4*) (Yilmaz *et al.*, 2009) were also hit (the associated SNP distance ranged from 0 to 629.1 kb and averaged 187.9 kb).

A large-scale population with a multiple genetic background provides increased power for detecting the regulating components of flowering time

Together, population size and structure, marker density and target trait architecture determine the power of a GWAS. In our study, a triple GWAS of the three large-scale panels, which broadly cover both US and Chinese maize germplasm, provided multiple examinations of the genetic components of maize flowering time. We expected to observe complementary results among the panels. Indeed, we detected two major flowering time regions (simultaneously related to DA, DS, ASI, DACV and DSCV) encompassing a distance of 23.2–49.0 Mb (US-NAM) and 41.1–81.1 Mb (CN-NAM) on chromosome 9. Through GWAS, two neighboring candidates, a photoperiod pathway gene *CONZ1* (34.6 Mb on chromosome 9) (Miller *et al.*, 2008) and a *FT*-like gene *ZCN26* (76.3 Mb on chromosome 9), were also detected.

In contrast, between the regions of *CONZ1* and *ZCN26* (described above), another strong association signal was also detected (59.8 Mb on chromosome 9) in the Ames panel (with the highest BPP of 0.81). Neighboring this location was the *ARABIDOPSIS THALIANA HOMEOBOX1*-like (*ATH1*) gene *HB21* (which regulates GA biosynthesis and is related to flowering time) (Proveniers *et al.*, 2007). Other examples of differences among the panels were association signals detected for three clustered candidates (GRMZM2G336909, *TCPTF8* and *BHLH32*) located within an intensively studied QTL region of 156–162 Mb on chromosome 3 (Buckler *et al.*, 2009). The central candidate, *TCPTF8* (*AtTCP20*-like gene) (Li *et al.*, 2005), was detected in both CN-NAM (519 kb upstream, with a BPP of 0.86) and Ames (937 kb downstream, with a BPP of 0.97). In US-NAM, GRMZM2G336909 contained a SNP with a BPP of 0.48. In Ames, *BHLH32* was neighbored by a SNP with a

BPP of 0.75 (503 kb upstream). Thus, *TCPTF8* might be the most reliable candidate in this region, but the two neighboring candidates may also affect flowering time in the different panels.

Moreover, we also observed the existence of several regions with strong associations, but without distributions of candidates or homologs. Such regions included 81–84 Mb on chromosome 1 (with the highest BPP of 0.68 in Ames), 31–40 Mb on chromosome 3 (with the highest BPP of 0.53 in US-NAM), 64–69 Mb on chromosome 4 (with the highest BPP of 0.82 in US-NAM), 142–143 Mb on chromosome 5 (with the highest BPP of 0.54 in CN-NAM) and 80–82 Mb on chromosome 6 (with the highest BPP of 0.81 in Ames). This finding suggests that these regions might contain additional unique genetic components or variants in the GRN of maize flowering time, which presumably are currently unknown.

In general, the genome-wide identification of 90 flowering time regions discloses a type of parallelism for the proposed model of numerous, additive small-effect QTLs that determine maize flowering time (Buckler *et al.*, 2009). Importantly, for those 220 candidates hit within a 1-Mb region of the associated SNPs, most appeared in the form of multi-homologs, such as the *PEBP*-like, *CO*-like, *SBP*-like and MADS-box genes. This finding reveals an added complexity to maize flowering time regulation, but is meaningful to the broad local adaptation of diverse maize germplasm.

However, the detection of marker–trait associations mainly depends on linkage disequilibrium (LD) between marker polymorphism and functional mutants. Detection is also influenced by other factors, including marker density, marker polymorphism information content (PIC), genetic background and the GWAS model. Therefore, in the future, complete sequence variants from the pan-genome of a large-scale panel (Tettelin *et al.*, 2005) – combined with multi-temporal–spatial expression profile information of diverse germplasms (obtained through RNA sequencing) – will advance the complete dissection of the maize flowering time GRN and other complex quantitative traits to improvements in food production.

EXPERIMENTAL PROCEDURES

Plant materials and phenotyping

Three maize panels, including two NAM population panels and one natural association panel, were used in this study. One of the NAM population panels was developed by Cornell University in the United States (US-NAM) and includes 5000 RILs from crosses of 25 diverse inbred lines: B97, CML52, CML69, CML103, CML228, CML247, CML277, CML322, CML333, Hp301, Il14H, Ki3, Ki11, Ky21, M37W, M162W, Mo18W, MS71, NC350, NC358, Oh43, Oh7B, P39, Tx303 and Tzi8, with the common parent 'B73' (Yu *et al.*, 2008). The other NAM population panel was developed by the Chinese Academy of Agricultural Sciences (CN-NAM) and includes 1971 RILs from 11 crosses of 11 diverse inbred lines: Zheng58, Ye478, Qi319, Weifeng322, Lv28, Pa405, Duo229, K12, Mo17, Huobai and Huangyesi3, with the common parent 'HZS' (Li *et al.*, 2013). To capture more genetic diversity, a natural association panel (Ames) with 1745 inbred lines from the USDA-ARS NCRPIS was also used (Romay *et al.*, 2013).

US-NAM has been phenotyped under eight environments in the United States (Buckler *et al.*, 2009). The flowering times of 4763 US-NAM lines were recorded at the following five additional locations in China in 2010 and 2011: (i) Hainan (18.15°N, 109.30°E) (10HN), (ii) Tongnan in Chongqing Province (30.03°N, 106.22°E) (11CQ), (iii) Xinxiang in Henan Province (35.19°N, 113.53°E) (11XX), (iv) Tianjin (39.40°N, 117.05°E) (11TJ), and (v) Beijing (39.48°N, 116.28°E) (11BJ). CN-NAM was phenotyped at the following three locations in China in 2009 and 2010, as described by Li *et al.*, 2013: (i) Xinxiang in Henan Province (35.19°N, 113.53°E) (09XX and 10XX), (ii) Beijing (39.48°N, 116.28°E) (09BJ and 10BJ), and (iii) Urumqi in Xinjiang Province (43.47°N, 87.39°E) (09XJ and 10XJ). In contrast, the Ames panel was repeatedly observed at Beijing, China (39.48°N, 116.28°E) in 2011 (11BJ).

The DA, DS and ASI were measured or calculated as described by Buckler *et al.* (2009). The broad sense heritability (h^2) for DA, DS and ASI was calculated according to Buckler *et al.* (2009) and Chandler *et al.* (2013). The BLUPs of DA, DS and ASI for both CN-NAM (six environments) and US-NAM (13 environments) were calculated using SAS PROC MIXED, with genotype, environment and replication as random effects (Brown *et al.*, 2011). The CVs of both the DA (DACV) and DS (DSCV) for CN-NAM across six environments (three locations across 2 years) and US-NAM across 13 environments were calculated to reflect the environmental sensitivity of these two flowering time traits. These values were calculated as CV (%) = SD/mean, where SD and mean refer to the standard deviation and mean of the DA or DS values across all of the environments for each RIL.

The QTL mapping of CN-NAM and US-NAM

For the single-RIL-family QTL mapping of the two NAMs, the ICIM method was used using the QTL ICI Mapping software v.3.2 (Wang, 2012). We set the *P* threshold at 0.001 for entering and 0.002 for removing. The LOD threshold was determined via a 1000-permutation test. Joint stepwise regressions for all five flowering time traits for both CN-NAM and US-NAM were conducted separately using SAS PROC GLMSELECT (Brown *et al.*, 2011). The method used stepwise regression with criteria of selection/removal thresholds set to $P = 1 \times 10^{-4}$. The thresholds for the entry and exit of the model terms were determined by permutation as follows: phenotypic data were randomly permuted within each family. All marker-by-family terms were tested and the lowest *P*-value was recorded for each permutation. A total of 1000 permutations were performed. After the model was fitted with stepwise regression, each marker was dropped from the model one at a time and a single best marker was refitted. This process tested the fitness of each marker using the remaining QTLs as background and also improved the overall fit of the model to the data (Buckler *et al.*, 2009).

High-density markers provide the opportunity to distinguish neighboring QTLs within a short physical distance. Meanwhile, the general strategy of a one-dimensional genome scan has been adopted for QTL mapping, which fits all possible single-QTL models but tends to ignore the reality of a possible tight-linked locus. To overcome this limitation, we use the two-dimensional examination approach (Broman and Sen, 2009) to test whether a short

chromosome region (<10 Mb) with a high LOD value in JSR should be treated as one functional unit or neighboring multi-units. In this process, the most significant marker within the short region is first treated as the covariant (marker effect nested by family) to calculate the contributions of its neighboring markers. If significant contributions (LOD ≥ 2.5) are demonstrated by its neighboring markers, we can hypothesize that more than one functional unit should be contained in this region. Then, the newly identified QTL marker (with the highest contribution among the tested markers) is treated as the covariate to test the contributions of the associated significant marker and its neighbors. In this manner, the most significant marker near the previous one is identified. We repeated these steps until the steady QTL markers were identified within the target region. The contributions of the neighboring markers were calculated using the following equation: $SS_{pop*mi} = SS_{total} - (SS_{pop} + SS_{pop*m} + SS_{mi\_residual})$, where SS_{pop*mi} is the contribution of the ith tested marker nested family, SS_{total} is the sum of the total squared contributions, SS_{pop} is the squared contribution of the family, SS_{pop*m} is the squared contribution of the most significant marker nested family and $SS_{mi\_residual}$ is the square of the residual when both the most significant marker and its ith neighboring marker are fitted to the model.

After all QTL markers had been identified, confidence intervals were estimated by sequentially examining flanking markers as described by Poland *et al.* (2011). That is, the full confidence interval was estimated using the flanking markers on both sides of a QTL marker. First, the QTL marker and the flanking marker from one side were fitted into the full linear model. The flanking marker was considered within the 95% confidence interval when the QTL marker lacked a significant contribution to the model ($P < 0.05$). The test was repeated until the QTL marker made a significant contribution. This procedure was then repeated for the flanking marker on the other side of the QTL marker.

Genome-wide association studies

A compressed mixed linear model (Zhang *et al.*, 2010) was conducted in the GAPIT R package (Lipka *et al.*, 2012) to test the marker–trait associations of each flowering time trait for each panel. To control false positives, a subsampling-based multiple SNP model was applied (Valdar *et al.*, 2006; Tian *et al.*, 2011). Briefly, 80% of the original entries were sampled in the new subpopulation, without replacement, and used to fit the SNPs with permutation-derived significance thresholds. This process was repeated 100 times. The BPP was then calculated as the proportion of the 100 replicates in which a trait-associated SNP was detected (ranging from 0 to 1). By using the 1000 times permutation test, we found that a threshold BPP of 0.05 led to a type I error rate < 0.05 and a BPP of 0.20 led to a type I error rate < 0.01.

Preliminarily, flowering time-associated SNPs detected in 5% or more of the replicates (BPP ≥ 0.05) were examined as polymorphisms in LD with potential candidate genes. To strictly control false discoveries, we first selected all SNPs with BPP ≥ 0.20. Meanwhile, some regions (less than 5 Mb) contained the SNPs that had their highest BPP values <0.20, but ≥0.05. These SNPs were also likely be neighbored by flowering time-related functional units with minor effects. Therefore, to avoid missing genetic components with weak association signals within these regions, a combined strategy was applied – the SNPs within a 5-Mb window were also selected if their highest BPP values were less than 0.20. Finally, these two sets of markers comprised the flowering time-associated SNPs.

Overlapping test between associated SNPs and QTLs

We evaluated the overlaps between the results of GWAS and linkage mapping by comparing locations between the flowering time-associated SNPs and the QTLs detected in the CN-NAM and US-NAM panels. If an associated SNP fell within a QTL interval, we deduced that this SNP overlapped with the relevant QTL; otherwise, we assumed no overlap. Then, we calculated the actual intra-QTL ratio (proportion of overlapped-associated SNPs) as the number of associated SNPs falling within QTLs divided by the total number of associated SNPs. In addition, different subsets of randomly selected SNPs (340 for CN-NAM, 367 for US-NAM) were sampled from 0.95 million SNPs across the genome, with 1000 replications. The proportion of these randomly selected SNPs falling within the QTL intervals was calculated as described above and defined as the expected (by chance) intra-QTL ratio.

Flowering time candidate gene enrichment tests

We hypothesized that the flowering time-associated SNPs should be closely linked to related causal variants. The tests for significant enrichments followed the description of Brown *et al.* (2011). The number of associated SNPs for all five flowering time traits that fell within different regions (0, 1, 2, 3, 4, 5, 10, 20, 30, 100, 300 and 500 kb, 1, 2 and 5 Mb) of candidate genes was compared with a null distribution. The null distribution was obtained by selecting an equivalent number of random genes and calculating their proximity to the detected associated SNPs. Then, the enrichment ratio was calculated by dividing the number of associated SNPs that fell within a certain region by the number of randomly selected genes that fell within that same region. This process was repeated 1000 times.

ACKNOWLEDGEMENTS

We thank Junjie Fu for sharing information about the maize gene list and annotations, Dallas Kroon for assisting with phenotype data management and Alex Lipka and Huihui Li for assisting with data analysis. This work was supported by the Ministry of Science and Technology of China (2011DFA30450, 2014BAD01B00, 2011CB100100, 2014DFG31690), NSF of China (91335206, U1138304), the 'Innovation Program' of CAAS, NSF #0820619 and #1238014, and the USDA-ARS of the United States of America, an Emerging Research Issues Internal Competitive Grant from the Agricultural Research Center at Washington State University, the Endowment and Research Project (no. 126593) from the Washington Grain Commission. We also thank Linda R. Klein for valuable advice and editing the manuscript.

CONFLICT OF INTEREST

The authors declare that they have no conflict of interest.

SUPPORTING INFORMATION

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

Figure S1. Flowering time quantitative trait locus distributions of all 36 recombinant inbred line families by inclusive composite interval mapping and joint linkage mapping for CN-NAM and US-NAM.

Figure S2. The mapping results for days to anthesis, days to silking, anthesis–silking interval, coefficient of variation (CV) of DA and CV of DAS by the method of joint stepwise regression for both CN-NAM and US-NAM.

Figure S3. Two neighboring flowering time quantitative trait loci on chromosome 8 can be distinguished by a two-dimensional examination approach.

Figure S4. Cross-validation of flowering time-associated single-nucleotide polymorphisms among CN-NAM, US-NAM and Ames for the trait of days to silking (five-fold, 100 replicates).

Figure S5. Scatter plot of principal component 1 (PC1) versus PC2 for 1515 Ames lines within the genome-wide association (GWAS) panel and those 835 lines outside (independent) of the GWAS panel.

Figure S6. The 243 associated single-nucleotide polymorphisms of Ames can predict the growing degree days days to silking of 434 genetically related lines outside (independent) of the Ames panel very well, but were unable to predict the phenotypic value of the 401 distantly related tropical lines.

Data S1. Materials used in this study and their flowering time observations in China.

Data S2. US-NAM recombinant inbred line family means for each flowering time trait.

Data S3. CN-NAM recombinant inbred line family means for each flowering time trait.

Data S4. Flowering time quantitative trait locus collection of all 36 recombinant inbred line families for the traits of days to anthesis (DA), days to silking (DS), anthesis–silking interval, coefficient of variation (CV) of DA and CV of DS.

Data S5. Flowering time regions identified by joint linkage analysis of CN-NAM and US-NAM.

Data S6. Flowering time-associated SNPs (BPP $\geq$ 0.05) for CN-NAM, US-NAM and Ames identified by genome-wide association studies.

Data S7. List of maize flowering time candidate genes or homologs.

Data S8. The 220 maize flowering time candidate genes or homologs within a 1-Mb region of associated single nucleotide polymorphisms.

REFERENCES

Andersen, C.H., Jesen, C.S. and Petersen, K. (2004) Similar genetic switch systems might integrate the floral inductive pathways in dicots and monocots. *Trends Plant Sci.* **9**, 105–107.

Brachi, B., Faure, N., Horton, M. *et al.* (2010) Linkage and association mapping of *Arabidopsis thaliana* flowering time in nature. *PLoS Genet.* **6**, e1000940.

Broman, K.W. and Sen, S. (eds) (2009) Two-dimensional, two-QTL scans. In *A guide QTL mapping with R/qtl*. New York: Springer, pp: 213-239.

Brown, P.J., Upadyayula, N., Mahone, G.S. *et al.* (2011) Distinct genetic architectures for male and female inflorescence traits of maize. *PLoS Genet.* **7**, e1002383.

Buckler, E.S., Holland, J.B., Bradbury, P.J. *et al.* (2009) The genetic architecture of maize flowering time. *Science*, **325**, 714–718.

Chandler, K., Lipka, A.E., Owens, B.E. *et al.* (2013) Genetic analysis of visually scored oranges kernel color in maize. *Crop Sci.* **53**, 714–718.

Chardon, F., Virton, B., Moreau, L. *et al.* (2004) Genetic architecture of flowering time in maize as inferred from quantitative trait loci meta–analysis and synteny conservation with the rice genome. *Genetics*, **168**, 2160–2185.

Chen, C., DeClerck, G., Tian, F., Spooner, W., McCouch, S. and Bucler, E.S. (2012) PICARA, an analytical pipeline providing probabilistic inference about *a priori* candidates genes underlying genome–wide association QTL in plants. *PLoS ONE*, **7**, e46596.

Chen, Q., Zhong, H., Fan, X.W. and Li, Y.Z. (2014) An insight into the sensitivity of maize to photoperiod changes under controlled conditions. *Plant, Cell Environ.* **38**, 1479–1489.

Colasanti, J. and Muszynski, M. (2009) The maize floral transition. In *Handbook of Maize: Its Biology* (Bennetzen, J. and Hake, S., eds). New York: Springer, pp 41–55.

Danilevskaya, O.N., Meng, X., Hou, Z., Ananiev, E.V. and Simmons, C.R. (2008) A genomic and expression compendium of the expanded *PEBP* gene family from maize. *Plant Physiol.* **146**, 250–264.

Dong, Z., Danileyskaya, O., Abadie, T., Messina, C., Cotes, N. and Cooper, M. (2012) A gene regulatory network model for floral transition of the shoot apex in maize and its dynamic modeling. *PLoS ONE*, **8**, e43450.

Ehrenreich, I.M., Hanzawa, Y., Chou, L., Roe, J.L., Kover, P.X. and Purugganan, M.D. (2009) Candidate gene association mapping of *Arabidopsis* flowering time. *Genetics*, **183**, 325–335.

Elzinga, J.A., Atlan, A., Biere, A., Gigord, L., Weis, A.E. and Bernasconi, G. (2007) Time after time: flowering phenology and biotic interactions. *Trends Ecol. Evol.* **22**, 432–439.

Endelman, J. (2013) Ridge regression and other kernels for genomic selection. Version 4.2.

Fornara, F., Panigrahi, K.C., Gissot, L. *et al.* (2010) SnapShot: control of flowering in *Arabidopsis*. *Cell*, **141**, 550–550.

Hung, H.Y., Shannon, L.M., Tian, F. *et al.* (2012) *ZmCCT* and the genetic basis of day–length adaptation underlying the postdomestication spread of maize. *Proc. Natl Acad. Sci. USA*, **109**, E1913–E1921.

Izawa, T., Takahashi, Y. and Yano, M. (2003) Comparative biology comes into bloom: genomic and genetic comparison of flowering pathways in rice and *Arabidopsis*. *Curr. Opin. Plant Biol.* **6**, 113–120.

Johnson, R. (2004) Marker–assisted selection. *Plant Breed. Rev.* **24**, 293–309.

Jung, C. and Müller, A. (2009) Flowering time control and applications in plant breeding. *Trends Plant Sci.* **14**, 563–573.

Kim, Y., Lim, J., Yeom, M. *et al.* (2013) ELF4 regulates GIGANTEA chromatin access through subnuclear sequestration. *Cell Rep.* **3**, 671–677.

Kong, F., Liu, B., Xia, Z. *et al.* (2010) Two coordinately regulated homologs of *FLOWERING LOCUS T* are involved in the control of photoperiodic flowering in soybean. *Plant Physiol.* **154**, 1220–1231.

Li, C., Potuschak, T., Colón-Carmona, A., Gutiérrez, R.A. and Doerner, P. (2005) *Arabidopsis* TCP20 links regulation of growth and cell division control pathways. *Proc. Natl Acad. Sci. USA*, **102**, 12978–12983.

Li, C.H., Li, Y.X., Sun, B.C. *et al.* (2013) Quantitative trait loci mapping for yield components and kernel–related traits in multiple connected RIL populations in maize. *Euphytica*, **193**, 303–316.

Li, C.H., Li, Y.X., Bradbury, P.J. *et al.* (2015) Construction of high-quality recombination maps with low-coverage genomic sequencing for joint linkage analysis in maize. *BMC Biol.* **13**, 78.

Lipka, A.E., Tian, F., Wang, Q. *et al.* (2012) GAPIT: genome association and prediction integrated tool. *Bioinformatics*, **28**, 2397–2399.

Malcomber, S.T., Prestoin, J.C., Reinheim, R. *et al.* (2006) Developmental gene evolution and the origin of grass inflorescence diversity. *Adv. Bot. Res.* **44**, 425–481.

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

Meng, X., Muszynski, M.G. and Danilevskaya, O.N. (2011) The *FT*–like *ZCN8* gene functions as a floral activator and is involved in photoperiod sensitivity in maize. *Plant Cell*, **23**, 942–960.

Miller, T., Muslin, E.H. and Dorweiler, J.E. (2008) A maize *CONSTANS*–like gene, *conz1*, exhibits distinct diurnal expression patterns in varied photoperiods. *Planta*, **227**, 1377–1388.

Osnato, M., Castillejo, C., Matías-Hernández, L. and Pelaz, S. (2012) TEMPRANILLO genes link photoperiod and gibberellin pathways to control flowering in *Arabidopsis*. *Nat. Commun.* **3**, 808.

Pin, P.A. and Nilsson, O. (2012) The multifaceted roles of *FLOWERING LOCUS T* in plant development. *Plant, Cell Environ.* **35**, 1742–1755.

Poland, J.A., Bradbury, P.J., Buckler, E.S. and Nelson, R.J. (2011) Genome-wide nested association mapping of quantitative resistance to northern leaf blight in maize. *Proc. Natl Acad. Sci. USA*, **17**, 6893–6898.

Proveniers, M., Rutjens, B., Brand, M. and Smeekens, S. (2007) The *Arabidopsis TALE* homeobox gene *ATH1* controls floral competency through positive regulation of FLC. *Plant J.* **52**, 899–913.

Rodgers-Melnick, E., Bradbury, P.J., Elshire, R.J. *et al.* (2015) Recombination in diverse maize is stable, predictable and associated with genetic load. *Proc. Natl Acad. Sci. USA*, **12**, 3823–3828.

Romay, M.C., Millard, M.J., Glaubitz, J.C. *et al.* (2013) Comprehensive genotyping of the USA national maize inbred seed bank. *Genome Biol.* **14**, R55. doi: 10.1186/gb-2013-14-6-r55.

Ryu, C.H., Lee, S., Cho, L.H. *et al.* (2009) OsMADS50 and OsMADS56 function antagonistically in regulating long day (LD)–dependent flowering in rice. *Plant, Cell Environ.* **32**, 1412–1427.

Salvi, S., Sponza, G., Morgante, M. *et al.* (2007) Conserved noncoding genomic sequences associated with a flowering time quantitative trait locus in maize. *Proc. Natl Acad. Sci. USA*, **104**, 11376–11381.

Schaub, M.A., Boyle, A.P., Kundaje, A., Batzoglou, S. and Snyder, M. (2012) Linking disease associations with regulatory information in the hunman genome. *Genome Res.* **22**, 1748–1759.

Song, Y.H., Lee, I., Lee, S.Y., Imaizumi, T. and Hong, J.C. (2012) CONSTANS and ASYMMETRIC LEAVES 1 complex is involved in the induction of FLOWERING LOCUS T in photoperiodic flowering in *Arabidopsis*. *Plant J.* **69**, 332–342.

Srikanth, A. and Schmid, M. (2011) Regulation of flowering time: all roads lead to Rome. *Cell. Mol. Life Sci.* **68**, 2013–2037.

Teng, W.T., Cao, J.S., Chen, Y.H. *et al.* (2004) Analysis of maize heterotic groups and patterns during past decade in China. *Sci. Agric. Sinica.* **37**, 1804–1811.

Tettelin, H., Masignani, V., Cieslewicz, M.J. *et al.* (2005) Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: implications for the microbial 'pan–genome'. *Science*, **102**, 13950–13955.

Thompson, B.E. (2009) Beyond the ABCs: MADS–box regulation of maize floral development. Maize Genetics Conference Abstracts. 51, T26.

Tian, F., Bradbury, P.J., Brown, P.J. *et al.* (2011) Genome–wide association study of leaf architecture in the maize nested association mapping population. *Nat. Genet.* **43**, 159–162.

Valdar, W., Solberg, L.C., Gauguier, D. *et al.* (2006) Genome–wide genetic association of complex traits in heterogeneous stock mice. *Nat. Genet.* **38**, 879–887.

Wallance, J.G., Bradbury, P.J., Zhang, N.Y., Gibon, Y., Stitt, M. and Buckler, E.S. (2015) Association mapping across numerous traits reveals Patterns of functional variation in maize. *PLoS Genet.* **10**, e1004845–e1004845.

Wang, J. (2012) *QTL IciMapping Software*. Beijing, China: Isbreeding.

Wang, R., Yu, Y., Zhao, J., Shi, Y., Song, Y., Wang, T. and Li, Y. (2008) Population structure and linkage disequilibrium of a mini core set of maize inbred lines in China. *Theor. Appl. Genet.* **117**, 1141–1153.

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

Wright, S.I., Bi, I.V., Schroeder, S.G., Yamasaki, M., Doebley, J.F., McMullen, M.D. and Gaut, B.S. (2005) The effects of artificial selection on the maize genome. *Science*, **308**, 1310–1314.

Wu, G., Park, M.Y., Conway, S.R., Wang, J.W., Weigel, D. and Poethig, R.S. (2009) The sequential action of miR156 and miR172 regulates developmental timing in *Arabidopsis*. *Cell*, **138**, 750–759.

Xue, W., Xing, Y., Weng, X. *et al.* (2008) Natural variation in *Ghd7* is an important regulator of heading date and yield potential in rice. *Nat. Genet.* **40**, 761–767.

Yang, Q., Li, Z., Li, W. *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.

Yilmaz, A., Nishiyama, M.Y. Jr, Fuentes, B.G. *et al.* (2009) GRASSIUS: a platform for comparative regulatory genomics across the grasses. *Plant Physiol.* **149**, 171–180.

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.

Zhang, Z.W., Ersoz, E., Lai, C.Q. *et al.* (2010) Mixed linear model approach adapted for genome-wide association studies. *Nat. Genet.* **42**, 355–360.