Genic and non-genic contributions to natural variation of quantitative traits in maize

Jianming Yu, Xianran Li, Chengsong Zhu, et al.

Genome Res. published online June 14, 2012
Access the most recent version at doi:10.1101/gr.140277.112

Genic and non-genic contributions to natural variation of quantitative traits in maize

Xianran Li[1], Chengsong Zhu[1], Cheng-Ting Yeh[2], Wei Wu[2], Elizabeth M. Takacs[3], Katherine A. Petsch[4], Feng Tian[5], Guihua Bai[1,6], Edward S. Buckler[5,6], Gary J. Muehlbauer[7], Marja C.P. Timmermans[4], Michael J. Scanlon[3], Patrick S. Schnable[2,*] & Jianming Yu[1,*]

[1] Department of Agronomy; Kansas State University; Manhattan, KS 66506, USA
[2] Center for Plant Genomics and Department of Agronomy; Iowa State University; Ames, IA 50011, USA
[3] Department of Plant Biology; Cornell University; Ithaca, NY 14853, USA
[4] Cold Spring Harbor Laboratory; Cold Spring Harbor, NY 11724, USA
[5] Institute for Genomic Diversity; Cornell University; Ithaca, NY 14853, USA
[6] United States Department of Agriculture-Agricultural Research Service (USDA-ARS)
[7] Department of Agronomy and Plant Genetics; University of Minnesota; St. Paul, MN 55108, USA

* Correspondence should be addressed to J.Y. (jyu@ksu.edu) or P.S.S. (schnable@iastate.edu)

Running title: Genic and non-genic contributions to natural variation

Abstract

The complex genomes of many economically important crops present tremendous challenges to understand the genetic control of many quantitative traits with great importance in crop production, adaptation, and evolution. Advances in genomic technology need to be integrated with strategic genetic design and novel perspectives to break new ground. Complementary to individual-gene targeted research, which remains challenging, a global assessment of the genomic distribution of trait-associated SNPs (TASs) discovered from genome scans of quantitative traits can provide insights into the genetic architecture and contribute to the design of future studies. Here we report the first systematic tabulation of the relative contribution of different genomic regions to quantitative trait variation in maize. We found that TASs were enriched in the non-genic regions, particularly within a 5 kb window upstream of genes, which highlights the importance of polymorphisms regulating gene expression in shaping the natural variation. Consistent with these findings, TASs collectively explained 44~59% of the total phenotypic variation across maize quantitative traits, and on average, 79% of the explained variation could be attributed to TASs located in genes or within 5 kb upstream of genes, which together comprise only 13% of the genome. Our findings suggest efficient, cost-effective GWAS in species with complex genomes can focus on genic and promoter regions.

Introduction

Cloning of individual large-effect genes underlying qualitative and quantitative traits has provided some insights into the genetic control of trait variation. These studies have most frequently implicated nucleotide polymorphisms in genic regions as being causative (Doebley et al. 2006; Miura et al. 2011); however, generalizing from these results may not be appropriate because of ascertainment bias, *e.g.*, preference given to genic regions during the efforts in gene cloning. In addition, it remains challenging to identify, validate, and characterize genes underlying modest-to-small-effect QTLs, which are common contributors to quantitative traits in crops with complex genomes. With these challenges, a different approach is to identify what part of a complex genome can be prioritized. Intuitively, this is known from mutational dissection of "qualitative phenotypes", but a global assessment for "quantitative traits" is lacking and the relative importance of genic and non-genic portions of the genome has significant bearings on further biological research and crop improvement.

By identifying trait-associated SNPs (TASs), genome-wide association studies (GWAS) can enhance our understanding of the genetic architecture (Chang et al. 2009; Meyer et al. 2008; Teslovich et al. 2010). For example, a survey of 531 human TASs found that most are located in non-coding regions (43% from non-genic regions and 45% from introns), suggesting that the search for functional polymorphisms should extend beyond coding regions (Hindorff et al. 2009). Indeed, some recent individual-gene studies have suggested that functional non-genic polymorphisms can also contribute to the variation associated with quantitative traits in plants (Ashikari et al. 2005; Clark et al. 2006; Frary et al. 2000; Salvi et al. 2007; Stam et al. 2002). However, previous GWAS in plants focused on single SNP testing or multiple-regression (Atwell et al. 2010; Huang et al. 2010; Tian et al. 2011) and did not address this critical issue. Hence, a systematic evaluation of TASs in plants can help to answer several important questions: (1) *What are the overall contributions of genetic polymorphisms (i.e., SNPs) in explaining the phenotypic variation of quantitative traits?* (2) *What are the relative contributions of genic and non-genic polymorphisms?* (3) *What is the distribution of maize TASs across different genomic annotation sets (e.g., promoter, intron, or coding region).*

Here we report genome scans of five quantitative traits with SNPs identified by two complementary next generation sequencing strategies to identify the underlying TASs, the genomic distribution of these TASs, and the relative contributions of genic and non-genic TASs to the phenotypic variation. We found that genic and non-genic TASs contribute approximately equally to phenotypic variation of maize quantitative traits. But the distributions of maize TASs in specific annotation sets differed. Specifically, nonsynonymous SNPs are under-represented among TASs for maize quantitative traits, suggesting that regulatory variation plays an important role in phenotypic variation. Our results suggest that genotyping methods designed to discover SNPs in genes and their upstream regions can be an economical approach for detecting genome-wide association signals in future GWAS scans of quantitative traits in crops with complex genomes.

Results

To be consistent with the GWAS literature (Hindorff et al. 2009), a genic region in this study is defined as between the transcription start site (TSS) and the end of 3' UTR. Toward this end, we first conducted RNA-seq to identify gene-enriched SNPs. A targeted-dissection genome scan

method was implemented to identify the TASs for five maize quantitative traits (leaf length, leaf width, upper leaf angle, days to anthesis, and days to silking) from 1 million SNPs merged from the RNA-seq data and the previously defined set of maize HapMap SNPs (Gore et al. 2009). We then systematically characterized the genomic distribution and genetic features of the discovered TASs across regions and traits. Two annotations sets (high stringent Filtered Gene Set, FGS; and low stringent Working Gene Set, WGS) have been released for maize genome (Schnable et al. 2009). Because the results obtained from the analysis of the FGS and WGS were similar, for simplicity of discussion only the FGS analysis results are presented in the text, but the results from the WGS analysis are included in the supplementary materials.

RNA-seq and SNPs for GWAS

RNA was extracted from shoot apices, which includes the shoot apical meristem (SAM) and up to five young leaf primordia, of two-week-old seedlings from each of the NAM founders (Methods) and used to conduct RNA-seq. To discover SNPs, the RNA-seq reads were aligned to the B73 maize reference genome sequence (Schnable et al. 2009). This tissue and stage of development were selected because we had previously shown that a substantial percentage of genes is expressed in the SAM and leaf primordia at this stage of development (Emrich et al. 2007), and because we were interested in testing whether SNPs identified from genes expressed at this stage of development are enriched for genes related to leaf architecture traits versus flowering-time traits.

Nearly a million SNPs (N = 942,793) were identified from ~600 million 76-bp RNA-seq reads from the 26 inbred founders (not including B73) of the NAM population. Of all discovered SNPs, only 289,461 that could be called with high confidence (Methods) in more than 81% of the NAM founders (*i.e.*, >22/27) were used for GWAS (**Table 1**). As expected based on their mode of discovery from RNA-seq reads, most of these SNPs were located within annotated genes Of the 289,461 SNPs, 87% were located within 15,097 of the 32,540 FGS genes (46%).

Similarly, we retained 774,754 HapMap (Gore et al. 2009) SNPs using the same threshold for missing SNP genotypes among the NAM founders (*i.e.*, >22/27). About half (49%) of the HapMap SNPs resided within 25,738 genes (79% of the FGS), and 85% of these genes were expressed in leaves (Li et al. 2010). In total, after merging the two data sets just over a million (1.01 M) SNPs, of which 58% were located within 26,382 (81%) FGS genes, were available for GWAS (**Table 1**, **Supplementary Fig. 1**). These SNPs were projected from the NAM founders to the ~5,000 RILs based on the previous results from genotyping the NAM population with 1,106 tagging SNPs (tSNP; Methods) (McMullen et al. 2009; Tian et al. 2011).

GWAS of quantitative traits

Genome scans using the 1.01 M merged SNPs identified TASs underlying five quantitative traits (**Supplementary Fig. 2**). This was accomplished using a two-stage scan method made possible by the genetic design of NAM (Yu et al. 2008). The first stage was a low-resolution mapping using 1,106 tSNPs. This analysis identified 164 QTLs for the five quantitative traits; 44/164 of the QTL regions were co-localized by the adjacent or common tSNPs. The 164 target QTL regions covered about 67% of the maize genome, with an average size of 400 Mb for each trait. In the second stage of the genome scan, the 1.01 M SNPs, which had been projected from the NAM founders to the RILs, were tested for associations with the five traits. Target regions were

defined as the 3 tSNPs to the left and the 3 tSNPs right of the tSNP most strongly associated with each QTL in the first genome-wide QTL scan (**Supplementary Fig. 3**). For each target regions, on average 7,319 SNPs located among the 7 tSNPs were analyzed in this step. In terms of answering the critical aforementioned questions, this targeted scan is significantly different (see Discussion) from previous analysis which individual SNP search across a chromosome is conducted. Here, only the target regions were scanned for TASs to avoid false discoveries that might otherwise be introduced from nearby regions. Genome-wide polygenic effects were controlled by including other QTLs in the model during the scan. To control for multiple testing, a minimum false discovery rate, *Q* value (Storey and Tibshirani 2003), was calculated for each TAS. The vast majority of TASs (85%) had a *Q* value from 1.0×10^{-4} to 4.7×10^{-105} (**Supplementary Table 1**). After the genome scan, 40/44 of the co-localized QTL regions were dissected to independent TASs, demonstrating the high resolving power of this GWAS (**Supplementary Fig. 2**). The common TASs detected for other 4 co-localized QTL regions were not unexpected because they were detected for two flowering-time traits that have a higher phenotypic correlation than other pairs of traits. RNA-seq SNPs yielded the strongest signals in 16 of the 164 dissected regions (**Supplementary Fig. 1**). Meanwhile, of the genes implicated by the highest signals from HapMap SNPs, 58% also harbored RNA-seq SNPs with strong association signals, demonstrating the value of using RNA-seq SNPs for GWAS.

Distribution of TASs

Only ~6% of the maize genome is genic (Schnable et al. 2009). Knowledge of the relative proportion of TASs located in genic and non-genic regions would shed light on the relative contributions of these two regions to quantitative trait variation. To test whether the final identification of TASs was context-independent (*i.e.*, to confirm that identifying a non-genic TAS was not simply due to a high proportion of non-genic SNPs within the starting SNP set), the 164 dissected QTL regions were also separately scanned for TASs using only the RNA-seq SNPs or only the HapMap SNPs in addition to the merged SNP dataset (**Fig. 1**, **Supplementary Fig. 4**). For each analysis and for each QTL region, we identified the TAS with the highest signal and classified it as genic or non-genic. Summarizing the results across all 164 dissected QTL regions provided a genome-wide picture of the relative distributions of genic and non-genic TASs (**Fig. 1**). From the merged SNP set, non-genic TASs were identified for 46% of the dissected regions even though only a few (6%) regions were classified as predominantly non-genic (*i.e.*, containing a higher proportion of non-genic than genic SNPs). From the gene-enriched RNA-seq SNP set, non-genic TASs were detected for 16% of dissected QTL regions. When only the HapMap SNP set was analyzed, however, the classification of TASs obtained was not independent of the classification of the regions. A similar pattern was observed when the top 5 highest associated SNPs within each region were tabulated as TASs (**Supplementary Table 2**). Linkage disequilibrium analysis among the top 5 highest associated SNPs suggested that the identified genomic regions and the TASs are well supported (**Fig. 2**). For the majority of these QTLs, the strongest association signals were concentrated within a window of 500-1000 bp around the TASs. Moreover, genic or non-genic TASs were equally likely to be identified within the QTL regions with either large or small genetic effects, as shown by the well-mixed pattern of TASs when the QTL effects were sorted in descending order for each trait (**Fig. 1**, **Supplementary Fig. 4, Supplementary Table 2**). Because very few shared TASs (3%) were found among traits and the proportion of non-genic TASs was consistent across traits, our findings from these five traits

may represent a general feature of TASs underlying other quantitative traits in plants, or at least cross-pollinated crops that have been subjected to intense selection.

TASs in different genomic annotation sets

We further evaluated the distribution of TASs by testing for their enrichment in eight sets of genomic annotations (non-genic region, promoter 5kb, promoter 1kb, 5' UTR, synonymous site, nonsynonymous site, intron, and 3' UTR; **Table 1**). Because promoters are typically located upstream of transcription start sites (TSSs), for the purposes of this analysis, we defined 1 kb of the TSS as the promoter 1kb, and likewise 5 kb upstream of the TSS as promoter 5kb, respectively. Three annotation sets (non-genic region, promoter 5kb, and nonsynonymous) exhibited significantly different proportions of genic or non-genic TASs from the tested SNPs (**Fig. 3**). TASs were over-represented in non-genic (P = 0.026) and promoter 5kb regions (P = 0.041; **Supplementary Fig. 5**). Although 13% of the 1.01 M SNPs were nonsynonymous, only 4% of TASs were nonsynonymous SNPs, indicating that nonsynonymous polymorphisms are under-represented among maize TASs ($P = 2.7\times10^{-6}$).

Of 27 TASs within the promoter 5kb set, 19 were located upstream of genes for which duplication or single-copy has been characterized in a maize genome duplication study (Schnable et al. 2011). Interestingly, 12 of these TASs are located in the upstream of genes with duplicated copies (4,507) and 7 are located in the upstream of genes without duplication (9,867).

Variation explained by TASs

In the NAM population, 74~89% of the phenotypic variation for each of the five analyzed quantitative traits is explained by 28~37 QTLs. Similar to findings from other complex trait studies, only a small fraction of the variation can be captured by individual TASs; however, these TASs collectively explained 44~53% of the total phenotypic variation (**Fig. 4**). And this reduction in the amount of explained phenotypic variation by TASs compared with QTLs (equivalent to 56~67% of heritability by TASs) is much smaller than what has been reported in human GWAS summaries, which is typically less than 10% (Manolio et al. 2009). As expected from the genetic design, contribution from an individual TAS to the overall phenotypic variation is determined by the allele frequency and genetic effect (**Fig. 5**). A relatively large contribution from a TAS required a balanced allele frequency (*i.e.*, approaching to 0.5), a modest to large genetic effect size, or both.

Upon partitioning TASs into genic or non-genic groups, 46~63% of the explained variation (*i.e.*, 21~35% of the total phenotypic variation) is attributed to genic TASs across the five traits. Because maize TASs were found to be over-represented in upstream promoter regions, we also examined the expanded genic region (gene + 1 kb or 5 kb upstream region). By including upstream regions, the percentage of explained variation increased to 53~73% (genes + 1kb) or 67~91% (genes + 5kb) (**Fig. 4**). Taken together, while the larger of the expanded genic region account for only 13% of the maize genome, they comprise 71% (by count) of the identified TASs and contribute to 79% (by variation explained) of the phenotypic variation captured by all TASs.

Candidate genes implicated by TASs

A variety of gene ontology (GO) terms were over-represented among the TAS-implicated candidate genes, suggesting that complex networks shape these traits (**Supplementary Table 3**). For example, the terms "response to hormone stimulus", "protein transporter activity", and "regulation of transcription, DNA-dependent" were among the most significantly over-represented. Two genic SNPs (one intronic and one in the coding region) within *ligueless2* (*lg2*), which is required to form the ligule/auricle hinge between leaf blade and sheath (Harper and Freeling 1996; Walsh et al. 1998), had the highest association values with upper leaf angle. This result differs from a previous analysis (Tian et al. 2011), in which a SNP downstream of *lg2* was identified, although the association signals at all three sites were among the strongest in both analyses. Meanwhile, a TAS for leaf width were located 95kb upstream of *rough sheath1*(*rs1*), which is expressed in the base of initiating leaves and young leaf primordial (Schneeberger et al. 1995). Importantly, no other predicted maize genes are located between this TAS and the *rs1* gene. Other candidate genes implicated by TASs included *zea agamous5* (Mena et al. 1995) and a YABBY transcription factor (GRMZM2G102218) for days to silking. Two AP2 domain proteins (GRMZM2G129777 and GRMZM2G421033) showed associations for leaf length and days to silking, respectively (**Supplementary Table 1**). This maize YABBY transcription factor is a homolog of *CRABS CLAW* in Arabidopsis which has been shown to function in carpel and nectary development (Lee et al. 2005)

Expression analysis

Investigation of our new expression profile (Methods) and published transcriptome data (Li et al. 2010) revealed that almost all TAS-implicated candidate genes were dynamically expressed among 10 tissues (4 embryo developmental stages, 2 types of meristems, and 4 leaf gradient zones) (**Supplementary Fig. 6**). Indeed, the expression level of the leaf length-associated AP2 gene was higher than another AP2 gene related to flowering time (**Supplementary Fig. 7**). On average, TAS-implicated candidate genes associated with leaf width had higher expression levels than other genes that were expressed in leaves (**Supplementary Table 4**).

To provide further functional analysis of the TASs, we examined the expression patterns of TAS-implicated candidate genes in an RNA-seq comparison between a *leafbladeless1* (*lbl1*) mutant and its wild-type controls. The *lbl1* mutant affects a variety of leaf developmental processes (Nogueira et al. 2007; Timmermans et al. 1998). The finding that three candidate genes (GRMZM2G047129 for leaf length; GRMZM2G083812 and GRMZM2G417843 for leaf width) were differentially expressed between *lbl1* mutant and wild type under one or both backgrounds (**Supplementary Fig. 8**) provides additional functional support for the relevant TASs.

Discussion

Maize, with extensive morphological variation and genetic diversity, has been exploited as a model species in studies of quantitative and population genetics, selection theory, domestication, breeding methods, and molecular genetics. Recent maize research in association mapping (Yu et al. 2006; Zhang et al. 2010), nested association mapping (NAM) (Buckler et al. 2009; Yu et al. 2008), genome-wide selection (Bernardo and Yu 2007; Riedelsheimer et al. 2012), and NAM-GWAS scan (Tian et al. 2011) represented some new contributions to complex trait dissection and selection in plants, many of which are facing similar challenges in. The maize NAM population consists of 5,000 recombinant inbred lines (RILs) derived from crossing 26 diverse

founders to a common parent, B73. The NAM design combines the merits of both linkage and linkage disequilibrium mapping to detect molecular polymorphisms underlying quantitative traits (Buckler et al. 2009; Yu et al. 2008), and represents a logical framework for conducting genome scans of multiple quantitative traits to answer the aforementioned questions.

New angle into more important questions

In the current study, GWAS scans of five quantitative traits (three leaf and two flowering-time traits) were conducted to estimate the relative contributions of genic and non-genic genetic variants to phenotypic variation. Instead of searching for signals on a chromosome base (Tian et al. 2011), the new targeted-dissection method minimized the influence of other TAS-containing regions on the search. More importantly, this method also allowed us to tabulate the TAS findings across regions to systematically address the contribution of genic and non-genic polymorphisms to quantitative traits, which was not possible with the previous analysis. Specifically, two complementary angles were presented: number of TASs in each class, and the proportion of phenotypic variation explained by TASs in each class. Unlike a previous analysis (Tian et al. 2011), the current study included SNPs derived from the analysis of RNA-seq reads to enable a valid comparison of the relative contributions of genic versus non-genic variants, and included the analyses of two non-leaf traits (*i.e.* flowering time) to enable broader inferences.

Value of RNA-seq for SNP discovery

RNA-seq has been proposed as a fast and inexpensive genotyping method (Barbazuk et al. 2007; Cirulli et al. 2010; Cloonan et al. 2008). Considering that the total number of HapMap SNPs is three times that of the RNA-seq SNP set, a direct comparison between two genotyping approaches for association tests is inappropriate; however, we demonstrated the potential of using RNA-seq to discover SNPs that are suitable for GWAS. A total of 16 of the TASs were identified from the RNA-seq SNP set and 58% of the genes captured by the HapMap TASs also harbored strongly associated RNA-seq SNPs. Because much overlap between genes expressed in shoot apices and other tissues (Sekhon et al. 2011), SNPs discovered from RNA-seq conducted on RNA isolated from shoot apices can be used to map a variety of complex traits. In our study, of these 16 TASs, 9 were identified for the two flowering-time traits and 7 for the three leaf architecture traits. More importantly, having RNA-seq SNPs enabled us to compare genome scan results with different proportions of genic SNPs (RNA-seq, HapMap, and merged). These comparisons diminished the influence of the distribution of SNPs on TASs (*i.e.*, ascertainment bias), which therefore enabled us to evaluate the relative contributions of genic and non-genic polymorphisms on phenotypic variation of quantitative traits.

Distribution of maize TASs and candidate genes

A long-standing question in evolutionary biology is the nature of the genetic variation that controls variation in quantitative traits. That is, is the variation mainly driven by differences in protein sequence, or by differences in gene expression patterns (Alonso-Blanco et al. 2009; Clark et al. 2006; Doebley et al. 2006)? In human GWASs, nonsynonymous SNPs are over-represented among TASs (Hindorff et al. 2009), which is consistent with a major role for alterations in protein sequences (Stenson et al. 2009). In contrast, our data demonstrate that among maize TAS, nonsynonymous sites are significantly under-represented. This is true regardless of whether we analyze the FGS (which is predicted to include few false-positive genes, but to be missing ~1/3 of genes) or the WGS (which is expected to include all of maize gene space plus an estimated

~50,000 non-genic sequences) (**Supplementary Fig. 9-11**). Interestingly, recent research has demonstrated the functional outcomes of synonymous mutations (Plotkin and Kudla 2011; Waldman et al. 2011). Tabulation of cloned rice QTLs, which often have larger effects in the respective populations than the TASs identified in the current study, also indicated the importance of expression differences (Miura et al. 2011). In addition, no splicing site mutation or premature stop was found to be TASs. Because the NAM founders were selected to represent a broad range of genetic diversity, these results suggest that in maize, alterations in protein sequence were quantitatively less important during evolution and selection in defining the natural variation in quantitative traits than changes in gene regulation, even though this type of protein sequence change mutations are often found in genetic studies using individuals with extreme phenotypes. Consistent with this hypothesis, we found that SNPs from promoter regions (5kb upstream of the genes) and non-genic region as a whole were over-represented among TASs (in both the FGS and WGS analyses). These findings imply that long-distance regulatory elements and other forms of causative variants, like structural variations (Lai et al. 2010; Swanson-Wagner et al. 2010), probably also play important roles in the diverse morphology of maize (Clark et al. 2006; Salvi et al. 2007; Stam et al. 2002).

For TAS-implicated candidate genes, the over-representation of several GO terms and expression analysis among tissues provided supporting evidence of the dynamic nature of underlying processes of these quantitative traits. Follow-up studies could be prioritized for several of them, which were highlighted with additional support.

Explained variation and GWAS strategy

As has been observed in the human GWAS (Manolio et al. 2009), maize TASs also captured a portion of phenotypic variation. One potential source of unexplained variation in human diseases could be the action of rare variants having large effects. With the NAM design, the allele frequency captured within the non-B73 founders determines the actual frequency in the population, which affects the phenotypic variation explained. The percentage of variation explained by the TASs with the lowest possible minor allele frequency in the maize NAM population (2%, *i.e.*, the case where only one NAM founder has different allele than B73) ranged from 0 to 9% for these traits, demonstrating that the contribution of low-frequency alleles varied among quantitative traits (**Fig. 5**).

Interestingly, although TASs explained less phenotypic variation than QTLs, for all five traits TASs yielded a better model fit than QTLs as indicated by Bayesian Information Criterion (BIC) values. While QTL analysis in the first stage required many more model parameters to capture multiple alleles, the focus of the genome scan was to identify TASs, all of which were biallelic. Because of a higher penalty associated with the greater number of model parameters, the model fit was better for TASs than QTLs. Future studies may be able to explain more of the phenotypic variation and achieve a better model fit by considering haplotypes and functional predictions of multiple SNPs (Dickson et al. 2010; Singleton et al. 2010).

Efficient GWAS by including promoter regions

One of the main goals of this study was to evaluate the relative contributions of genic and non-genic polymorphisms to phenotypic variation for quantitative traits. Three relevant findings emerged from this study: (1) approximately half of the dissected regions had genic TASs and the

other half non-genic TASs, (2) genic or non-genic TASs contributed nearly half of phenotypic variation explained by all TASs, and (3) non-genic polymorphisms were significantly over-represented among TASs indicating the importance of analyzing non-genic regions in GWAS (Hindorff et al. 2009; Yang et al. 2011). But more importantly, because a significant fraction of the non-genic TASs and their corresponding contributions to the phenotypic variation explained were from the promoter regions upstream of genes (**Fig. 3-4**, **Supplementary Fig. 10-11**), targeting polymorphisms within genes and their promoter regions deserves attention.

GWAS across a very large number of samples genotyped by whole-genome sequencing would provide the most comprehensive understanding of genetic architecture. The cost of such an approach could, however, be prohibitive, especially for species with large genomes (Cirulli and Goldstein 2010). Our results suggest that at least in maize the bulk of TASs are located in genes and their promoters. Hence, the combination of RNA-seq and exome capture experiments using long-read (*e.g.*, 454) and paired-end (Illumina and 454) technologies would facilitate the cost-effective identification of promoter and genic polymorphisms for routine GWAS scans in species with complex genomes.

METHODS

Maize Population

The NAM population was developed by crossing 25 diverse founders (the NAM founders) to a common parent B73, whose genome serves as the reference genome for maize (Schnable et al. 2009). Approximately 200 recombination inbred lines (RILs) were developed from each cross (McMullen et al. 2009). All founders and 4,892 RILs (including 200 RILs from the intermated B73 × Mo17 population) were genotyped with 1,106 tagging SNPs (tSNPs). A consensus genetic map of the NAM population was constructed based on the tSNPs. The maize HapMap project sequenced genomic DNA from the NAM founders using an Illumina Genome Analyzer and discovered SNPs by aligning the reads to the RefGen_v1B73 genome (Gore et al. 2009).

Discovery of SNPs via RNA-seq

Total RNA was extracted from the shoot apex from two-week-old seedlings of the NAM founders using Trizol reagent (Invitrogen). The Poly(A) RNA was isolated from total RNA with oligo d(T) beads, then used to construct the RNA-seq libraries, which were sequenced using the Illumina Genome Analyzer II instrument.

Raw sequence reads were scanned for low quality bases. Nucleotides with PHRED quality values <15 were trimmed. Each read was examined in two phases. In the first phase, reads were scanned starting at each end and nucleotides with quality values lower than the threshold were removed. The remaining nucleotides were then scanned using overlapping windows of 10bp and sequences beyond the last window with average quality value less than the specified threshold were truncated. Trimmed reads were then aligned to the B73 reference genome (RefGen_v1) using GSNAP (Wu and Watanabe 2005) and uniquely mapped reads (≤2 mismatches every 36 bp and ≤3 bp tails allowed) were used for SNP discovery using the 123SNP software (available at http://schnablelab.plantgenomics.iastate.edu/software/). SNP sites were called only if the site contains at least 3 reads supporting the base call with error rates <0.03% and the most common allele must account for at least 80% of all aligned reads covering that nucleotide position.

SNP Annotation

SNPs were assigned as genic or non-genic based on their locations relative to two sets of annotated maize genes from maize RefGen_v1. The Filtered Gene Set (FGS, N = 32,540) includes stringently called genes and is therefore expected to include only a low number of false-positive calls, but to incompletely sample the full gene space. Based on an analysis of an independent RNA-seq data, nearly all (99%) of the 15,097 genes from the FGS that include SNPs are expressed in maize seedling leaves (Li et al. 2010). In contrast, the Working Gene Set (WGS, N = 109,563) was generated using less stringent parameters and is therefore expected to sample more of the gene space, but includes more false-positive gene calls. The current estimate of gene number in maize also considers the Classical Gene Set, a collection of N=464 well characterized ("true") genes, each of which has been cloned, confers a phenotype when mutated, and is supported by at least three citations (Schnable and Freeling 2011). Approximately 2/3 of the classical gene set is included within the FGS, suggesting that the true gene number in maize is ~50,000 (3/2 × 32,540). The entire classical gene set is included in the WGS, suggesting that the WGS contains most of the gene space.

For genes with alternative splicing isoforms, the transcript with the longest coding region was selected. The functional effects of SNPs (synonymous mutation, nonsynonymous mutation, splice site mutation, premature stop, or frame shift) were annotated using custom perl scripts.

Statistical analysis

To avoid over-prediction, SNPs for which genotyping data were missing for more than 5 of the NAM founders were not included in the association analysis. Other missing genotypes were imputed by fastPHASE 1.2 using default parameters (Scheet and Stephens 2006). The imputed SNPs between two adjacent informative tSNPs were projected from founders to RILs based on information from tSNPs (Tian et al. 2011).

The phenotypic data was available at Panzea database (www.panzea.org). In brief, three leaf traits (leaf length, leaf width, and upper leaf angle) and two flowering times (days to anthesis and days to silking) were collected in eight summer environments across two years. The best linear unbiased predictor (BLUP) of each line was used for linkage and association analysis. Joint linkage analysis across BLUP phenotype and 1,106 of tSNPs was conducted in SAS 9.1 with the GLMSelect Procedure. The parameters (population effect and tSNP by population effect) were estimated. tSNPs that met the significant *P* value (7.1×10^{-5}) of the marginal F-test in the final model were identified as QTLs.

GWAS was conducted in R (www.r-project.org). To control for polygenic background effects, we conducted the scan by including other QTLs located 10 cM away from the testing region (**Supplementary Fig. 4**). A dissected region was bordered by three flanking tSNPs from each side of a QTL tSNP (**Supplementary Fig. 4**). Each dissected region was classified as either predominantly genic if it contains a higher proportion of genic than non-genic SNPs, or predominantly non-genic if otherwise. The SNP with the highest association value was retrieved from each dissected QTL region and designated as the TAS. In cases where multiple TASs were tied in association signal strength, these TASs were individually classified and the ratio of genic:non-genic TASs determined the final classification (**Fig. 2**, **Supplementary Fig. 2**). Test of independence was conducted between the classifications of TASs and the target regions. To

address multiple testing issues, the *Q* value (*i.e.*, adjusted *P* value by adopting a false discovery rate controlling procedure) was estimated with Qvalue (Storey and Tibshirani 2003).

The likelihood-ratio-based R^2 was used for calculating the phenotypic variation explained by all QTLs and the same procedure for all TASs (Sun et al. 2010). The contributions from genic and non-genic TASs were calculated based on sum of squares from the combined model with both types of TASs fitted, and the same procedure for the contribution from TASs with the lowest minor-allele-frequency.

We mapped the SNPs (including TASs) into 8 annotation sets (non-genic region, promoter 5kb, promoter 1kb, intron, 3' UTR, 5' UTR, synonymous site, and nonsynonymous site). Non-genic region included promoter 5kb, which was defined as from the upstream 5kb to the TSS. Similarly, promoter 1kb covered the region from the upstream 1kb to the TSS. To avoid the overlap issue, the TAS distribution among different annotation used only the data from 114 QTL regions with unique TASs. The proportion of TASs in each annotation set was tested for enrichment or depletion against the corresponding proportion of SNPs.

Candidate gene

Gene ontology (GO) annotations of the filtered gene set from maize RefGen_v1 were obtained from MaizeSequence (www.maizesequence.org). Over-representation of GO terms with more one TAS-implicated gene was tested by comparing the proportion of the specific GO term among all implicated genes against all annotated maize genes. Significance of over-representation was corrected with the Bonferroni method (Rice 1989).

We sequenced RNA from four developmental stages of the embryo (proembryo, transition phase, coleoptile stage, and L1 stage embryo), and SAMs and lateral meristems (from two-week-old seedlings) of B73 (available at www.maizegdb.org). The expression patterns of TAS-implicated candidate genes were also analyzed using previously published transcriptome data from four zones (basal zone, transitional zone, maturing zone, and mature zone) along the proximodistal gradient of maize B73 seedling leaves (Li et al. 2010).

Transcriptome analysis of lbl1 mutants

Whole embryos of *lbl1* mutants and wild type (B73 or W22, into which *lbl1* had been introgressed) were isolated from developing seeds. mRNA from embryos was then extracted. RNA-seq was conducted on these samples using an Illumina Genome Analyzer II with three replications. The TAS-implicated candidate genes for leaf length and leaf width were screened to determine whether they were differentially expressed.

DATA ACCESS

The RNA-seq data from shoot apex have been deposited in the NCBI Sequence Read Archive with the reference number SRA050451.

ACKNOWLEDEGMENTS

This work is supported by the National Science Foundation (DBI-0820610).

AUTHOR CONTRIBUTIONS

J.Y., P.S.S., M.C.P.T., G.J.M, E.S.B., and M.J.S. designed the study. X.L. C.Z., C.-T.Y., W.W., E.M.T., K.A.P, and F.T. performed the analyses. X.L., J.Y., P.S.S., M.J.S., M.C.P.T., G.J.M., and G.B. drafted the manuscript. All authors critically revised and provided final approval of this manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

REFERENCES

Alonso-Blanco, C., M.G. Aarts, L. Bentsink, J.J. Keurentjes, M. Reymond, D. Vreugdenhil, and M. Koornneef. 2009. What has natural variation taught us about plant development, physiology, and adaptation? *Plant Cell* **21:** 1877-1896.

Ashikari, M., H. Sakakibara, S. Lin, T. Yamamoto, T. Takashi, A. Nishimura, E.R. Angeles, Q. Qian, H. Kitano, and M. Matsuoka. 2005. Cytokinin oxidase regulates rice grain production. *Science* **309:** 741-745.

Atwell, S., Y.S. Huang, B.J. Vilhjalmsson, G. Willems, M. Horton, Y. Li, D. Meng, A. Platt, A.M. Tarone, T.T. Hu et al. 2010. Genome-wide association study of 107 phenotypes in *Arabidopsis thaliana* inbred lines. *Nature* **465:** 627-631.

Barbazuk, W.B., S.J. Emrich, H.D. Chen, L. Li, and P.S. Schnable. 2007. SNP discovery via 454 transcriptome sequencing. *Plant J* **51:** 910-918.

Bernardo, R. and J. Yu. 2007. Prospects for genomewide selection for quantitative traits in maize. *Crop Science* **47:** 1082-1090.

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

Chang, B.L., S.D. Cramer, F. Wiklund, S.D. Isaacs, V.L. Stevens, J. Sun, S. Smith, K. Pruett, L.M. Romero, K.E. Wiley et al. 2009. Fine mapping association study and functional analysis implicate a SNP in *MSMB* at 10q11 as a causal variant for prostate cancer risk. *Hum Mol Genet* **18:** 1368-1375.

Cirulli, E.T. and D.B. Goldstein. 2010. Uncovering the roles of rare variants in common disease through whole-genome sequencing. *Nat Rev Genet* **11:** 415-425.

Cirulli, E.T., A. Singh, K.V. Shianna, D. Ge, J.P. Smith, J.M. Maia, E.L. Heinzen, J.J. Goedert, and D.B. Goldstein. 2010. Screening the human exome: a comparison of whole genome and whole transcriptome sequencing. *Genome Biol* **11:** R57.

Clark, R.M., T.N. Wagler, P. Quijada, and J. Doebley. 2006. A distant upstream enhancer at the maize domestication gene *tb1* has pleiotropic effects on plant and inflorescent architecture. *Nat Genet* **38:** 594-597.

Cloonan, N., A.R. Forrest, G. Kolle, B.B. Gardiner, G.J. Faulkner, M.K. Brown, D.F. Taylor, A.L. Steptoe, S. Wani, G. Bethel et al. 2008. Stem cell transcriptome profiling via massive-scale mRNA sequencing. *Nat Methods* **5:** 613-619.

Dickson, S.P., K. Wang, I. Krantz, H. Hakonarson, and D.B. Goldstein. 2010. Rare variants create synthetic genome-wide associations. *PLoS Biol* **8:** e1000294.

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

Emrich, S.J., W.B. Barbazuk, L. Li, and P.S. Schnable. 2007. Gene discovery and annotation using LCM-454 transcriptome sequencing. *Genome Res* **17:** 69-73.

Frary, A., T.C. Nesbitt, S. Grandillo, E. Knaap, B. Cong, J. Liu, J. Meller, R. Elber, K.B. Alpert, and S.D. Tanksley. 2000. fw2.2: a quantitative trait locus key to the evolution of tomato fruit size. *Science* **289:** 85-88.

Gore, M.A., J.M. Chia, R.J. Elshire, Q. Sun, E.S. Ersoz, B.L. Hurwitz, J.A. Peiffer, M.D. McMullen, G.S. Grills, J. Ross-Ibarra et al. 2009. A first-generation haplotype map of maize. *Science* **326:** 1115-1117.

Harper, L. and M. Freeling. 1996. Interactions of *liguleless1* and *liguleless2* function during ligule induction in maize. *Genetics* **144:** 1871-1882.

Hindorff, L.A., P. Sethupathy, H.A. Junkins, E.M. Ramos, J.P. Mehta, F.S. Collins, and T.A. Manolio. 2009. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc Natl Acad Sci U S A* **106:** 9362-9367.

Huang, X., X. Wei, T. Sang, Q. Zhao, Q. Feng, Y. Zhao, C. Li, C. Zhu, T. Lu, Z. Zhang et al. 2010. Genome-wide association studies of 14 agronomic traits in rice landraces. *Nat Genet* **42:** 961-967.

Lai, J., R. Li, X. Xu, W. Jin, M. Xu, H. Zhao, Z. Xiang, W. Song, K. Ying, M. Zhang et al. 2010. Genome-wide patterns of genetic variation among elite maize inbred lines. *Nat Genet* **42:** 1027-1030.

Lee, J.Y., S.F. Baum, J. Alvarez, A. Patel, D.H. Chitwood, and J.L. Bowman. 2005. Activation of *CRABS CLAW* in the nectaries and carpels of Arabidopsis. *Plant Cell* **17:** 25-36.

Li, P., L. Ponnala, N. Gandotra, L. Wang, Y. Si, S.L. Tausta, T.H. Kebrom, N. Provart, R. Patel, C.R. Myers et al. 2010. The developmental dynamics of the maize leaf transcriptome. *Nat Genet* **42:** 1060-1067.

Manolio, T.A., F.S. Collins, N.J. Cox, D.B. Goldstein, L.A. Hindorff, D.J. Hunter, M.I. McCarthy, E.M. Ramos, L.R. Cardon, A. Chakravarti et al. 2009. Finding the missing heritability of complex diseases. *Nature* **461:** 747-753.

McMullen, M.D., S. Kresovich, H.S. Villeda, P. Bradbury, H. Li, Q. Sun, S. Flint-Garcia, J. Thornsberry, C. Acharya, C. Bottoms et al. 2009. Genetic properties of the maize nested association mapping population. *Science* **325:** 737-740.

Mena, M., M.A. Mandel, D.R. Lerner, M.F. Yanofsky, and R.J. Schmidt. 1995. A characterization of the MADS-box gene family in maize. *Plant J* **8:** 845-854.

Meyer, K.B., A.T. Maia, M. O'Reilly, A.E. Teschendorff, S.F. Chin, C. Caldas, and B.A. Ponder. 2008. Allele-specific up-regulation of *FGFR2* increases susceptibility to breast cancer. *PLoS Biol* **6:** e108.

Miura, K., M. Ashikari, and M. Matsuoka. 2011. The role of QTLs in the breeding of high-yielding rice. *Trends Plant Sci* **16:** 319-326.

Nogueira, F.T., S. Madi, D.H. Chitwood, M.T. Juarez, and M.C. Timmermans. 2007. Two small regulatory RNAs establish opposing fates of a developmental axis. *Genes Dev* **21:** 750-755.

Plotkin, J.B. and G. Kudla. 2011. Synonymous but not the same: the causes and consequences of codon bias. *Nat Rev Genet* **12:** 32-42.

Rice, W.R. 1989. Analyzing tables of statistical tests. *Evolution* **43:** 223-225.

Riedelsheimer, C., A. Czedik-Eysenberg, C. Grieder, J. Lisec, F. Technow, R. Sulpice, T. Altmann, M. Stitt, L. Willmitzer, and A.E. Melchinger. 2012. Genomic and metabolic prediction of complex heterotic traits in hybrid maize. *Nat Genet* **44:** 217-220.

Salvi, S., G. Sponza, M. Morgante, D. Tomes, X. Niu, K.A. Fengler, R. Meeley, E.V. Ananiev, S. Svitashev, E. Bruggemann et al. 2007. Conserved noncoding genomic sequences associated with a flowering-time quantitative trait locus in maize. *Proc Natl Acad Sci U S A* **104:** 11376-11381.

Scheet, P. and M. Stephens. 2006. A fast and flexible statistical model for large-scale population genotype data: applications to inferring missing genotypes and haplotypic phase. *Am J Hum Genet* **78:** 629-644.

Schnable, J.C. and M. Freeling. 2011. Genes identified by visible mutant phenotypes show increased bias toward one of two subgenomes of maize. *PLoS One* **6:** e17855.

Schnable, J.C., N.M. Springer, and M. Freeling. 2011. Differentiation of the maize subgenomes by genome dominance and both ancient and ongoing gene loss. *Proc Natl Acad Sci U S A* **108:** 4069-4074.

Schnable, P.S. D. Ware R.S. Fulton J.C. Stein F. Wei S. Pasternak C. Liang J. Zhang L. Fulton T.A. Graves et al. 2009. The B73 maize genome: complexity, diversity, and dynamics. *Science* **326:** 1112-1115.

Schneeberger, R.G., P.W. Becraft, S. Hake, and M. Freeling. 1995. Ectopic expression of the knox homeo box gene *rough sheath1* alters cell fate in the maize leaf. *Genes Dev* **9:** 2292-2304.

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

Singleton, A.B., J. Hardy, B.J. Traynor, and H. Houlden. 2010. Towards a complete resolution of the genetic architecture of disease. *Trends Genet* **26:** 438-442.

Stam, M., C. Belele, W. Ramakrishna, J.E. Dorweiler, J.L. Bennetzen, and V.L. Chandler. 2002. The regulatory regions required for *B'* paramutation and expression are located far upstream of the maize *b1* transcribed sequences. *Genetics* **162:** 917-930.

Stenson, P.D., M. Mort, E.V. Ball, K. Howells, A.D. Phillips, N.S. Thomas, and D.N. Cooper. 2009. The Human Gene Mutation Database: 2008 update. *Genome Med* **1:** 13.

Storey, J.D. and R. Tibshirani. 2003. Statistical significance for genomewide studies. *Proc Natl Acad Sci U S A* **100:** 9440-9445.

Sun, G., C. Zhu, M.H. Kramer, S.S. Yang, W. Song, H.P. Piepho, and J. Yu. 2010. Variation explained in mixed-model association mapping. *Heredity* **105:** 333-340.

Swanson-Wagner, R.A., S.R. Eichten, S. Kumari, P. Tiffin, J.C. Stein, D. Ware, and N.M. Springer. 2010. Pervasive gene content variation and copy number variation in maize and its undomesticated progenitor. *Genome Res* **20:** 1689-1699.

Teslovich, T.M. K. Musunuru A.V. Smith A.C. Edmondson I.M. Stylianou M. Koseki J.P. Pirruccello S. Ripatti D.I. Chasman C.J. Willer et al. 2010. Biological, clinical and population relevance of 95 loci for blood lipids. *Nature* **466:** 707-713.

Tian, F., P.J. Bradbury, P.J. Brown, H. Hung, Q. Sun, S. Flint-Garcia, T.R. Rocheford, M.D. McMullen, J.B. Holland, and E.S. Buckler. 2011. Genome-wide association study of leaf architecture in the maize nested association mapping population. *Nat Genet* **43:** 159-162.

Timmermans, M.C., N.P. Schultes, J.P. Jankovsky, and T. Nelson. 1998. *Leafbladeless1* is required for dorsoventrality of lateral organs in maize. *Development* **125:** 2813-2823.

Waldman, Y.Y., T. Tuller, A. Keinan, and E. Ruppin. 2011. Selection for translation efficiency on synonymous polymorphisms in recent human evolution. *Genome Biol Evol* **3:** 749-761.

Walsh, J., C.A. Waters, and M. Freeling. 1998. The maize gene *liguleless2* encodes a basic leucine zipper protein involved in the establishment of the leaf blade-sheath boundary. *Genes Dev* **12:** 208-218.

Wu, T.D. and C.K. Watanabe. 2005. GMAP: a genomic mapping and alignment program for mRNA and EST sequences. *Bioinformatics* **21:** 1859-1875.

Yang, J., T.A. Manolio, L.R. Pasquale, E. Boerwinkle, N. Caporaso, J.M. Cunningham, M. de Andrade, B. Feenstra, E. Feingold, M.G. Hayes et al. 2011. Genome partitioning of genetic variation for complex traits using common SNPs. *Nat Genet* **43:** 519-525.

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

Yu, J., G. Pressoir, W.H. Briggs, I. Vroh Bi, M. Yamasaki, J.F. Doebley, M.D. McMullen, B.S. Gaut, D.M. Nielsen, J.B. Holland et al. 2006. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nat Genet* **38:** 203-208.

Zhang, Z., E. Ersoz, C.Q. Lai, R.J. Todhunter, H.K. Tiwari, M.A. Gore, P.J. Bradbury, J. Yu, D.K. Arnett, J.M. Ordovas et al. 2010. Mixed linear model approach adapted for genome-wide association studies. *Nat Genet* **42:** 355-360.

Table 1. Distribution of maize SNPs across different genomic annotation sets.

Dataset	Non-genic	Promoter 5kb	Promoter 1kb	Intron	5' UTR	3' UTR	Syn	Nsy
HapMap	394,160 (51%)	149,018 (19%)	77,513 (10%)	109,912 (14%)	47,312 (6%)	33,264 (4%)	98,434 (13%)	83,271 (11%)
RNA-seq	36,416 (13%)	8,863 (3%)	2,588 (1%)	5,663 (2%)	15,597 (5%)	62,983 (22%)	101,907 (35%)	66,355 (23%)
Merged	424,703 (42%)	156,197 (16%)	79,448 (8%)	114,784 (11%)	57,814 (6%)	88,152 (9%)	176,435 (18%)	134,484 (14%)

Non-genic includes promoter 5kb, which in turn includes promoter 1kb. UTR = untranslated region. Syn = synonymous SNPs. Nsy = nonsynonymous SNPs.

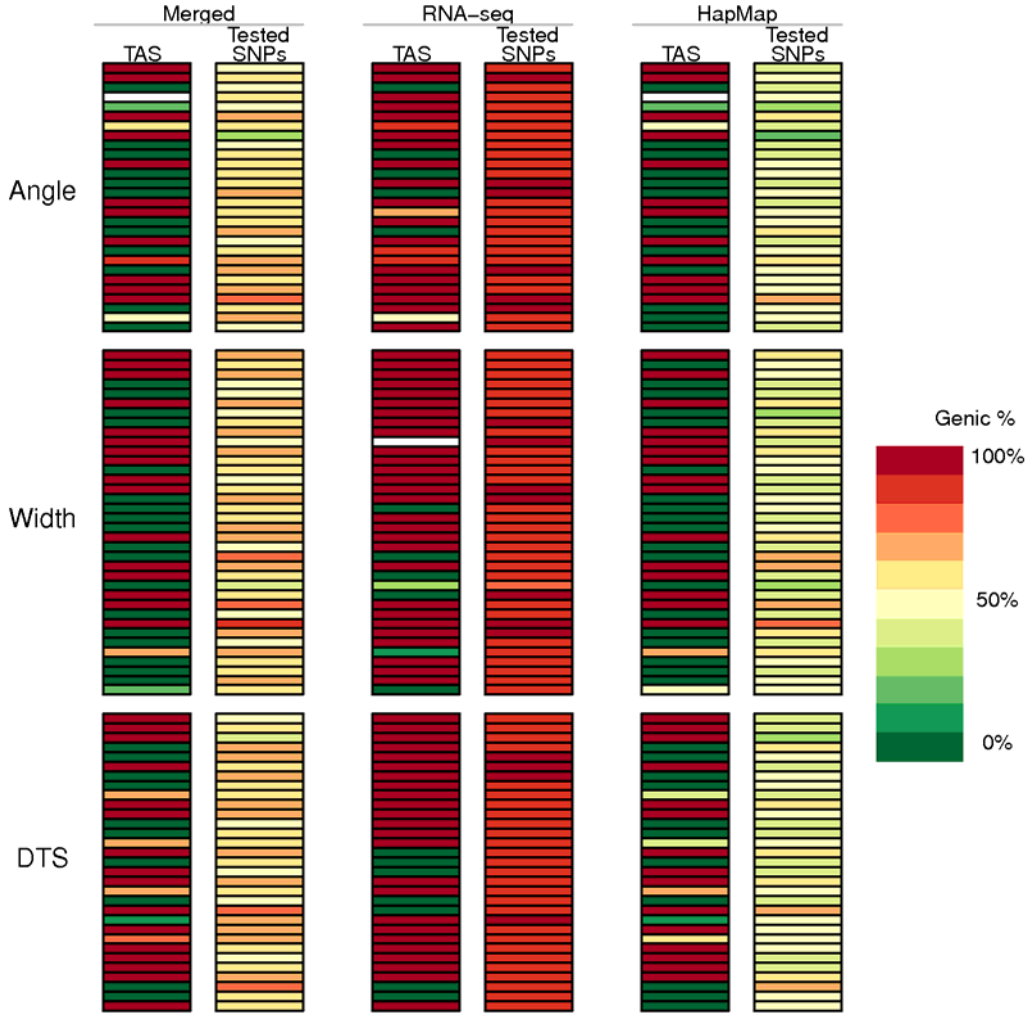


Figure 1. Genic or non-genic TASs for each dissected QTL region and the proportion of genic SNPs among all tested in the region. Genic region is defined as from the transcription start site to the end of 3' UTR. With the merged data set, the probability of a TAS being genic or non-genic is significantly independent of whether more genic or non-genic SNPs in the target region are tested ($P < 0.05$) for all five traits. Each row represents a single QTL region for the indicated traits. Genic and non-genic TASs are equally likely to be identified for QTLs with large or small effects, which are sorted in descending order within each trait. DTS, days to silking.

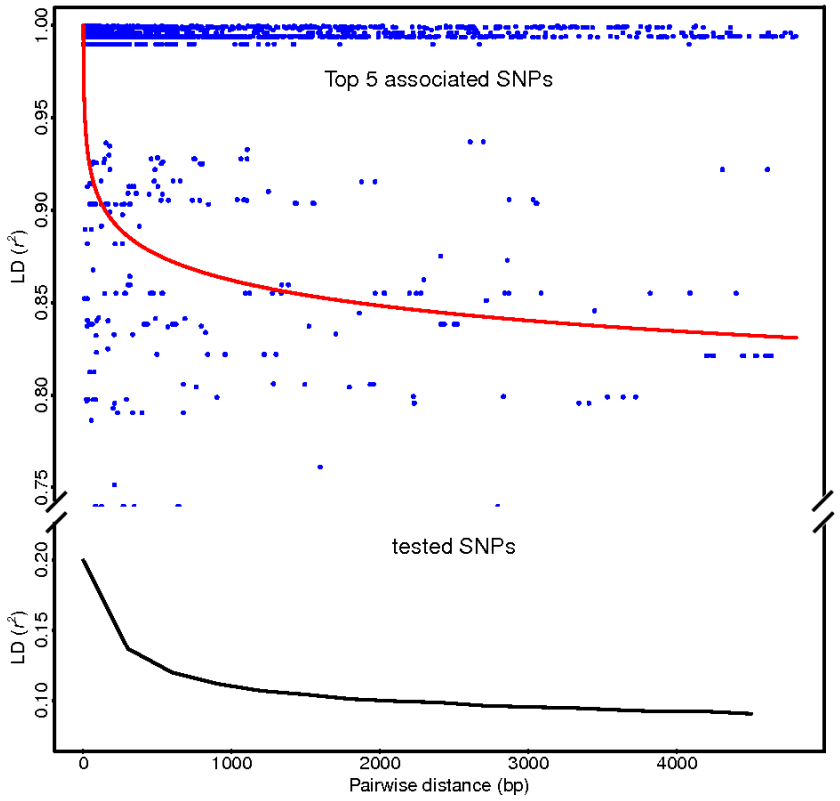


Figure 2. Linkage disequilibrium (LD) among Top 5 associated SNPs within each target region and the LD among all tested SNPs. Strong LD among Top 5 associated SNPs and plateau beyond 500-1,000 bp indicate genomic regions signaled by TASs are well supported.

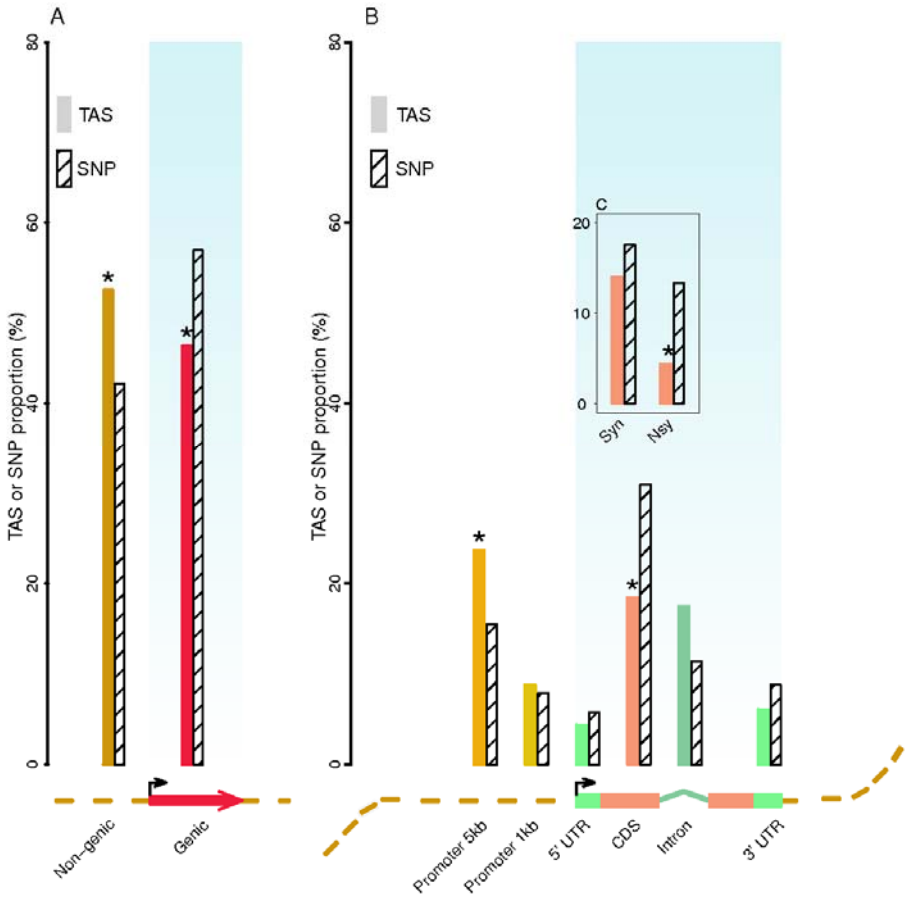


Figure 3. Distribution of TASs and tested SNPs for five quantitative traits across genomic annotation sets. (A) non-genic versus genic (genic region is defined as from the transcription start site to the end of 3' UTR); (B) different annotation sets; and (C) synonymous versus nonsynonymous. Genic plus 5-kb upstream regions comprise only 13% of the maize genome, but account for 71% of TASs. Non-genic and promoter 5kb regions are over-represented among TASs. The nonsynonymous set is under-represented among TASs. Note that non-genic region includes promoter 5kb, which in turn includes promoter 1kb, and genic region includes the untranslated region (UTR), coding region (CDS), and intron. Stars denote the proportion of TASs from the annotation set significantly differs from that of tested SNPs. Arrow denotes transcription start site.

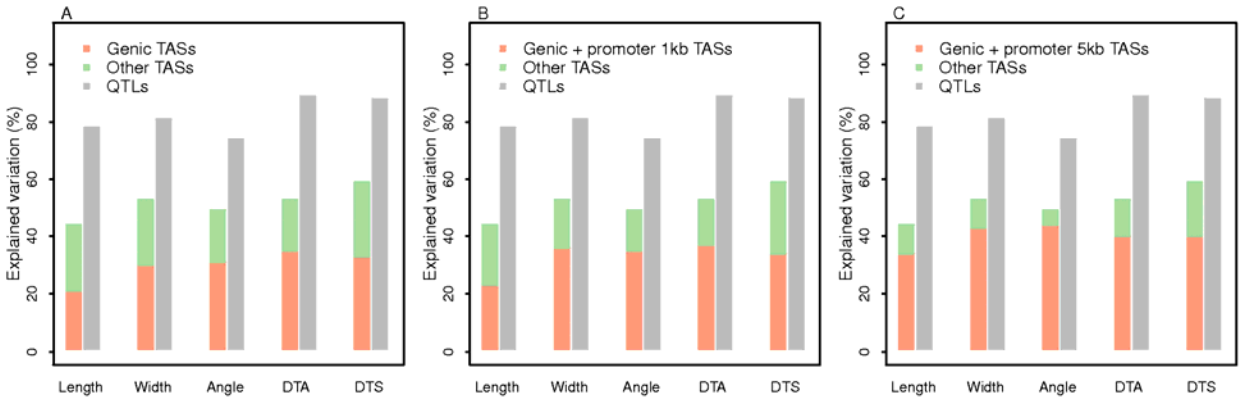


Figure 4. Phenotypic variation explained by genic and non-genic TASs. (A) genic TASs versus all other TASs, (B) genic region plus promoter 1kb versus all other TASs, and (C) genic region plus promoter 5kb versus all other TASs. Phenotypic variation explained by all QTLs is shown for comparison. TASs located within genic region plus upstream 5kb comprise only 13% of the maize genome, but explain a large proportion (67~91%) of the phenotypic variation captured by all TASs.

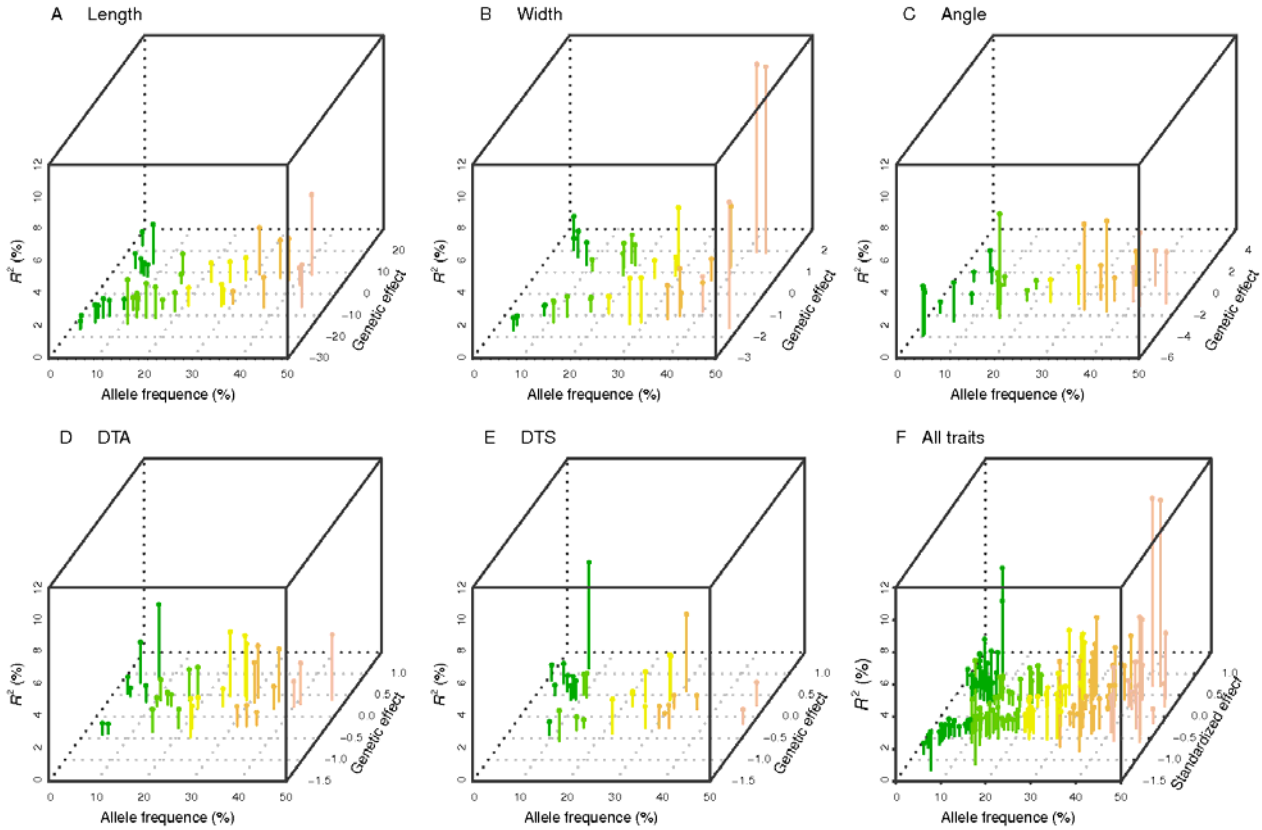


Figure 5. Contribution from an individual TAS to the phenotypic variation is determined by allele frequency and genetic effect size. (A), Leaf length, mm (B), Leaf width, mm; (C), Leaf angle, degree; (D), DTA, days to anthesis, day; (E), DTS, days to silking, day; and (F) Combined results for all traits with standardized genetics effects.

Supplementary materials

Supplementary Table 1. The genomic location, statistical significance, allele frequency, and implicated genes of the identified TASs underlying five maize quantitative traits from the merged SNP dataset.

Supplementary Table 2. Partition of the dissected QTL regions into regions with either predominantly genic or non-genic SNPs, and the identified TAS into genic or non-genic classes across five quantitative traits.

Trait	Region/TAS classification	HapMap			RNA-seq			merged		
		Tested SNP	Top 1 TAS	Top 5 TASs	Tested SNP	Top 1 TAS	Top 5 TASs	Tested SNP	Top 1 TAS	Top 5 TASs
Length	Genic	18	20	23	37	33*	36	36	20**	27**
	Non-genic	19	17	14	0	4*	1	1	17**	10**
Width	Genic	18	17	24	36	29**	35	33	17**	26**
	Non-genic	18	19	12	0	7**	1	3	19**	10**
Angle	Genic	9	13	10	28	23**	24	26	14**	15**
	Non-genic	19	14	18	0	5**	4	2	13**	13**
DTA	Genic	10	19**	14	32	28*	32	29	17**	22**
	Non-genic	22	13**	18	0	4*	0	3	15**	10**
DTS	Genic	12	17	13	31	24**	31	30	20**	20**
	Non-genic	19	14	18	0	7**	0	1	11**	11**

Star indicates whether the TAS is genic or non-genic (or whether more TASs coming from one class in "Top 5 TASs") is statistically independent of whether the dissected region contains more genic or non-genic SNPs. * $P < 0.05$; ** $P < 0.01$.

Supplementary Table 3. Over-represented GO terms of TAS-implicated genes. A variety of over-represented terms suggests that complex networks shape these quantitative traits.

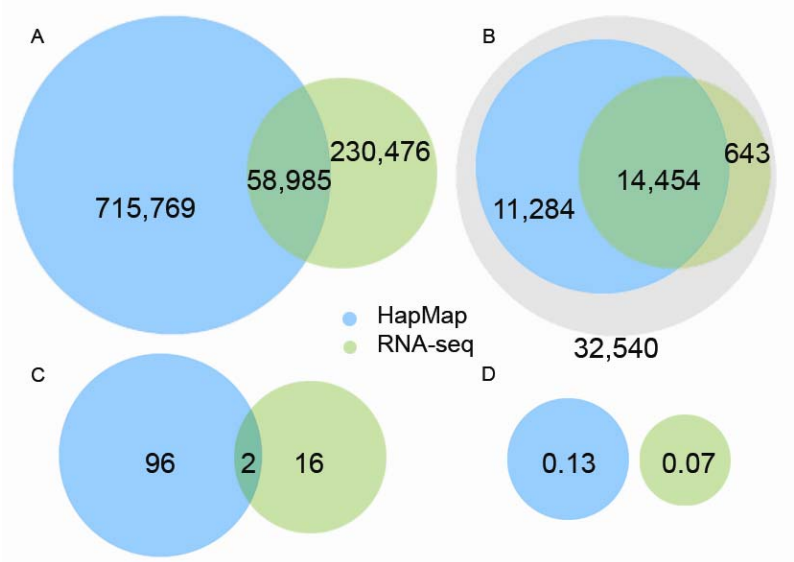
Trait	GO group	Go terms	Percentage among candidates	Percentage among genome	GO annotation
Length	Biological process	GO:0016192	3.409%	0.207%	Vesicle-mediated transport
		GO:0006886	3.409%	0.390%	Intracellular protein transport
		GO:0007165	2.273%	0.293%	Signal transduction
		GO:0015031	2.273%	0.333%	Protein transport
	Cellular component	GO:0030117	3.409%	0.065%	Membrane coat
	Molecular function	GO:0008565	2.273%	0.085%	Protein transporter activity
		GO:0016788	2.273%	0.155%	Hydrolase activity, acting on ester bonds
		GO:0003924	2.273%	0.271%	GTPase activity
		GO:0004672	5.682%	1.392%	Protein kinase activity
Width	Biological process	GO:0007242	4.167%	0.064%	Intracellular signaling cascade
		GO:0006413	4.167%	0.076%	Translational initiation
		GO:0006629	8.333%	0.360%	Lipid metabolic process
		GO:0009725	6.250%	0.053%	Response to hormone stimulus
		GO:0045449	6.250%	0.952%	Regulation of transcription
	Molecular function	GO:0003743	4.167%	0.081%	Translation initiation factor activity
		GO:0004629	4.167%	0.018%	Phospholipase C activity
		GO:0008081	8.333%	0.058%	Phosphoric diester hydrolase activity
Angle	Biological process	GO:0006355	9.677%	1.646%	Regulation of transcription, DNA-dependent
	Cellular component	GO:0005634	9.677%	2.656%	Nucleus
	Molecular function	GO:0046983	6.452%	0.601%	Protein dimerization activity
		GO:0043565	6.452%	0.763%	Sequence-specific DNA binding
DTA	Biological process	GO:0030001	3.333%	0.165%	Metal ion transport
		GO:0006355	6.667%	1.646%	Regulation of transcription, DNA-dependent
	Molecular function	GO:0046873	3.333%	0.040%	Metal ion transmembrane transporter activity
		GO:0046983	6.667%	0.601%	Protein dimerization activity
		GO:0043565	6.667%	0.763%	Sequence-specific DNA binding
		GO:0003700	6.667%	1.051%	Transcription factor activity
		GO:0003676	8.333%	2.049%	Nucleic acid binding
DTS	Biological process	GO:0009239	2.239%	0.096%	Enterobactin biosynthetic process
	Biological process	GO:0050826	7.463%	1.349%	Response to freezing
		GO:0042309	7.463%	1.349%	Homoiothermy
		GO:0008152	8.209%	2.274%	Metabolic process
		GO:0006915	1.493%	0.120%	Apoptosis
	Cellular component	GO:0005761	1.493%	0.012%	Mitochondrial ribosome
		GO:0005576	3.731%	0.338%	Extracellular region
		GO:0015935	1.493%	0.069%	Small ribosomal subunit
	Molecular function	GO:0004867	3.731%	0.092%	Serine-type endopeptidase inhibitor activity
		GO:0008667	2.239%	0.096%	2,3-dihydro-2,3-dihydroxybenzoate dehydrogenase activity
		GO:0016829	1.493%	0.069%	Lyase activity
		GO:0050825	7.463%	1.349%	Ice binding
		GO:0005488	5.970%	1.585%	Binding
		GO:0000287	1.493%	0.163%	Magnesium ion binding
		GO:0003824	6.716%	2.407%	Catalytic activity

Significant level is corrected with the Bonferroni method at $P < 0.05$. After the analysis of 80 TAS-implicated genes, only GO terms associated with more than one candidate genes were tabulated to avoid false-positive findings.

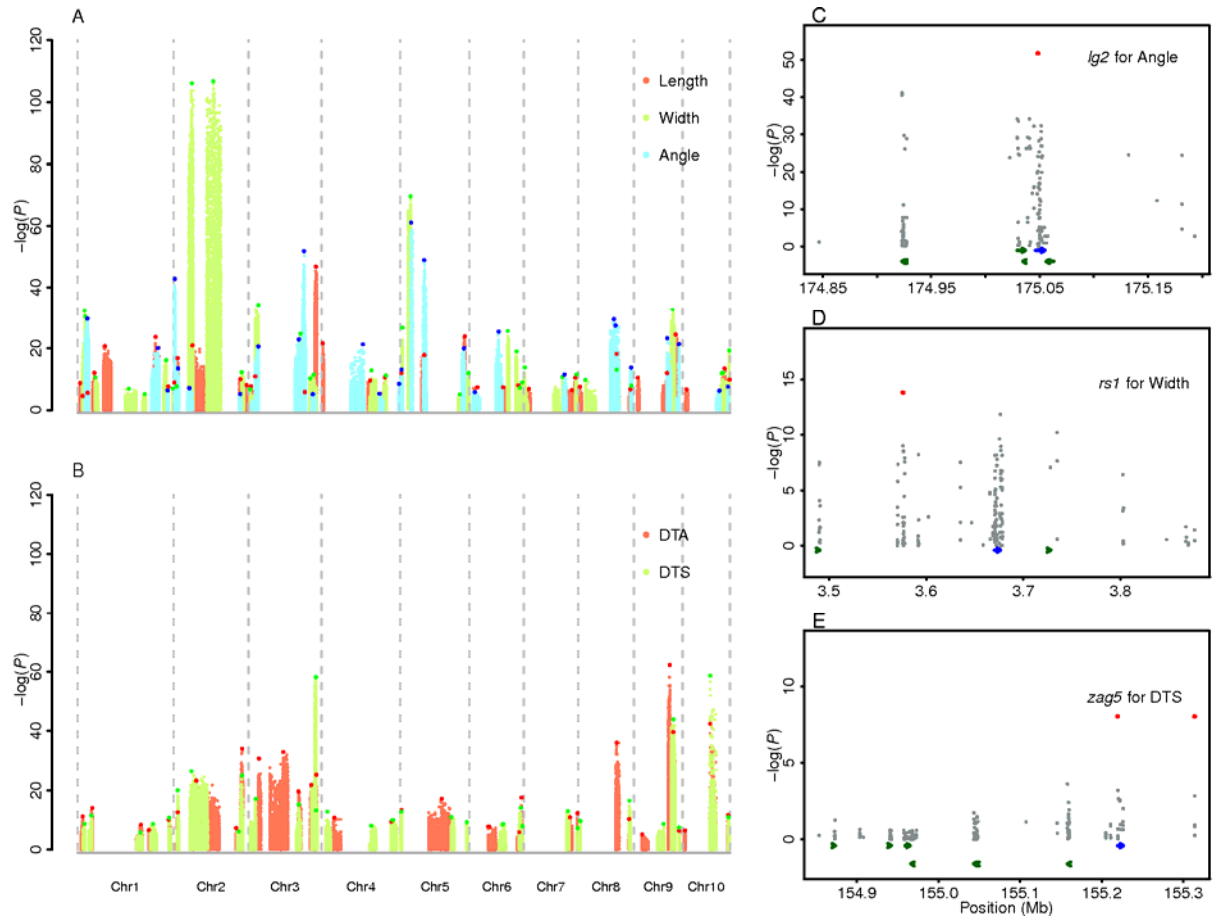
Supplementary Table 4. Expression level of TAS-implicated candidate genes among different tissues and leaf gradient zones.

	Embryo developmental stages				Meristem		Leaf gradient zones			
Trait	Proembryo	Transition phase	Coleoptile	L1 stage embryo	SAM	LM	Basal zone	Transitional zone	Maturing zone	Mature zone
Length	3.67	3.81	1.87	2.05	7.96	7.30	11.05	5.15	2.33	3.35
Width	6.31	2.50	3.05	2.75	4.38	2.60	12.31	9.38	5.98	7.49
Angle	4.15	5.80	2.93	2.58	4.32	3.72	6.87	2.01	1.65	1.70
Overall[1]	4.52	4.50	4.60	4.29	4.58	4.46	6.56	5.05	2.87	3.17

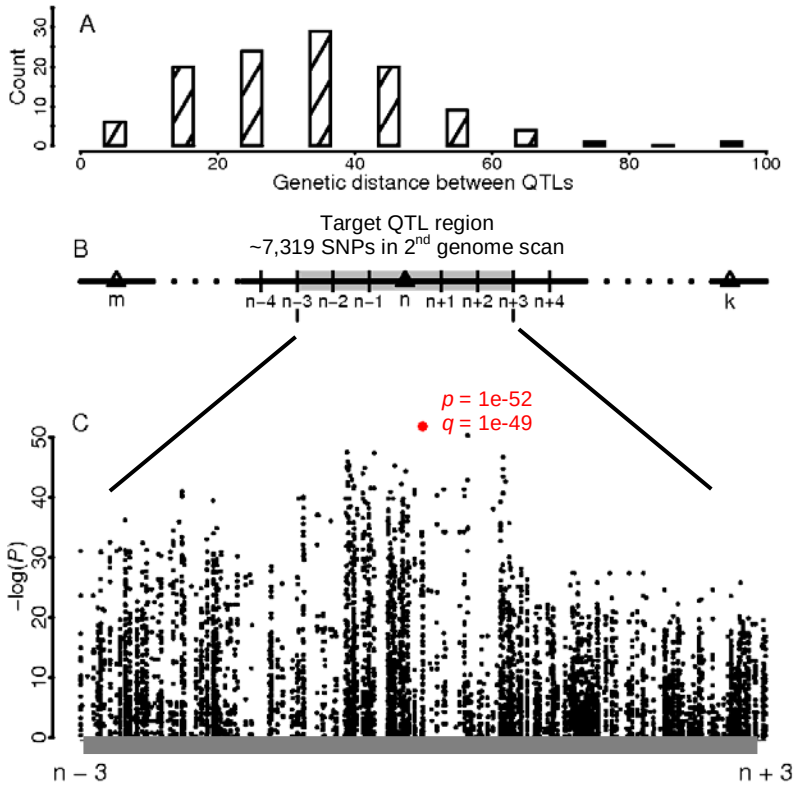
[1] Median value of all expressed genes within each tissue. SAM, shoot apical meristem from two-week-old seedling; LM, lateral meristem from two-week-old seedling. The number of each cell is the median value of RPKM.



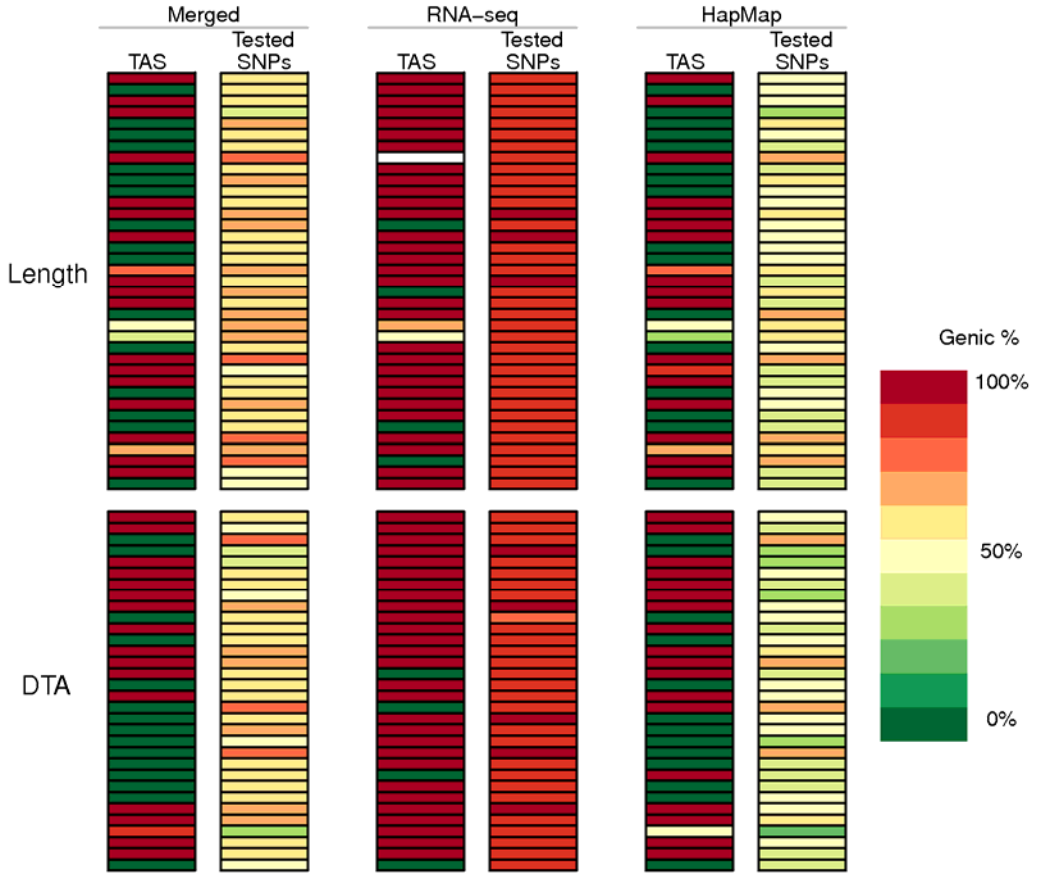
Supplementary Figure 1. Comparison between HapMap and RNA-seq for (A) Total number of SNPs, (B) Captured genes, (C) TASs, and (D) TAS per thousand SNPs. The filtered gene set (32,540) from maize B73 RefGen_v1 was analyzed.



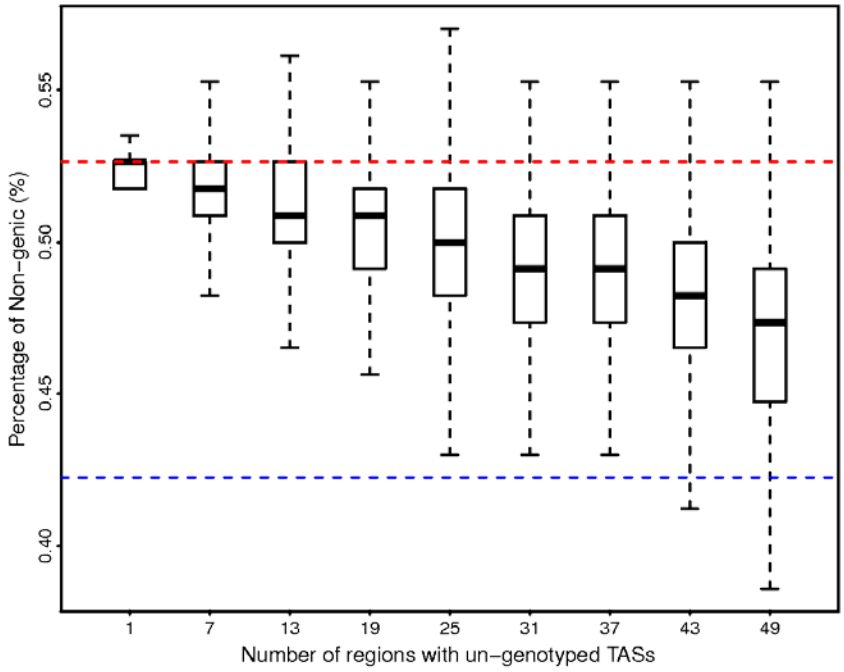
Supplementary Figure 2. GWAS for five quantitative traits with one million merged SNPs from the RNA-seq and HapMap data sets to identify TASs. (A), Three leaf architecture traits; and (B) Two flowering-time traits. Dark-colored dots indicate TASs with the highest association values for each of 164 QTL regions. TASs in three genes implicated for three traits are illustrated as examples to show TAS locations in (C), (D), and (E). Blue and green arrows indicate implicated and nearby genes, respectively. Red dots indicate TASs for tagged genes. DTA, days to anthesis; DTS, days to silking.



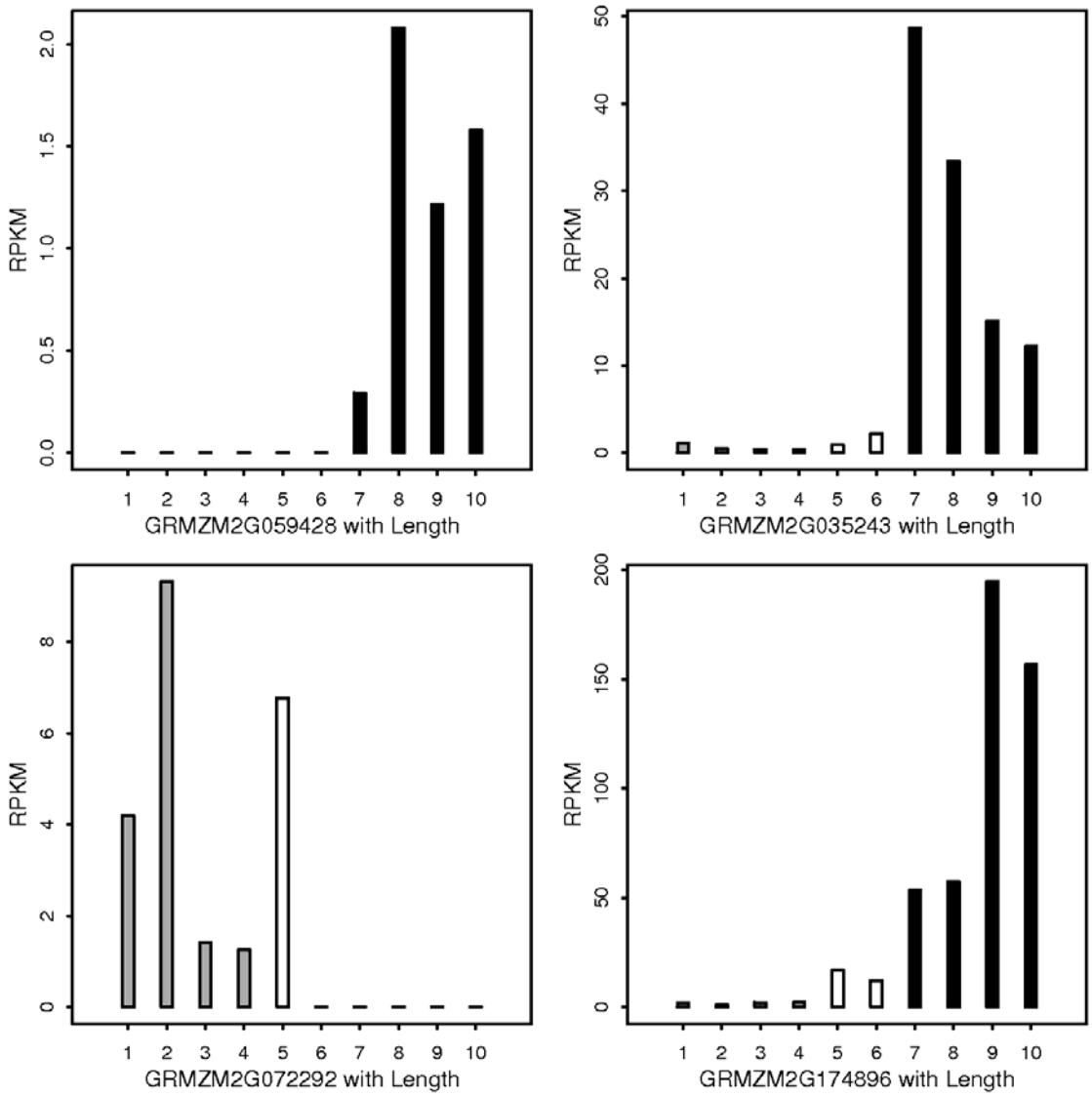
Supplementary Figure 3. Diagram of the target QTL region in the two-stage genome scan. (**A**) The distribution of genetic distance between two adjacent QTLs across five quantitative traits. Over 90% of adjacent QTLs had genetic distance larger than 10 cM. (**B**) Target region defined by three tagging SNPs to the left and right of the QTL *n*. On average, 7,319 SNPs were scanned in the second stage in each target region for TASs. The control of polygenic background effects during the scan is achieved by fitting all other QTLs (*e.g.*, *m* and *k*) located 10 cM away from the target region. (**C**) Detailed view of a TAS (red dot) within the target region. Both *P*-value and *Q*-value (adjusted for multiple testing) are indicated.



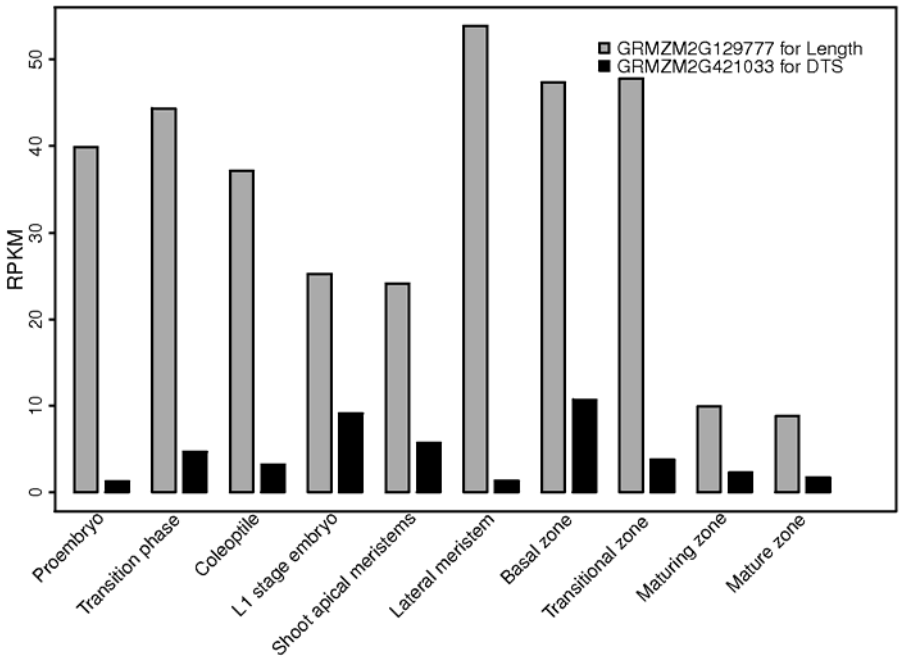
Supplementary Figure 4. Genic or non-genic TASs for each dissected QTL region and the proportion of genic SNPs among all tested in the region for Leaf Length and Days to Anthesis (DTA). Genic region is defined as from the transcription start site to the end of 3' UTR. With the merged data set, the probability of a TAS being genic or non-genic is significantly independent of whether more genic or non-genic SNPs in the target region are tested ($P < 0.05$) for all five traits. Each row represents a single QTL region for the indicated traits. Genic and non-genic TASs are equally likely to be identified for QTLs with large or small effects, which are sorted in descending order within each trait.



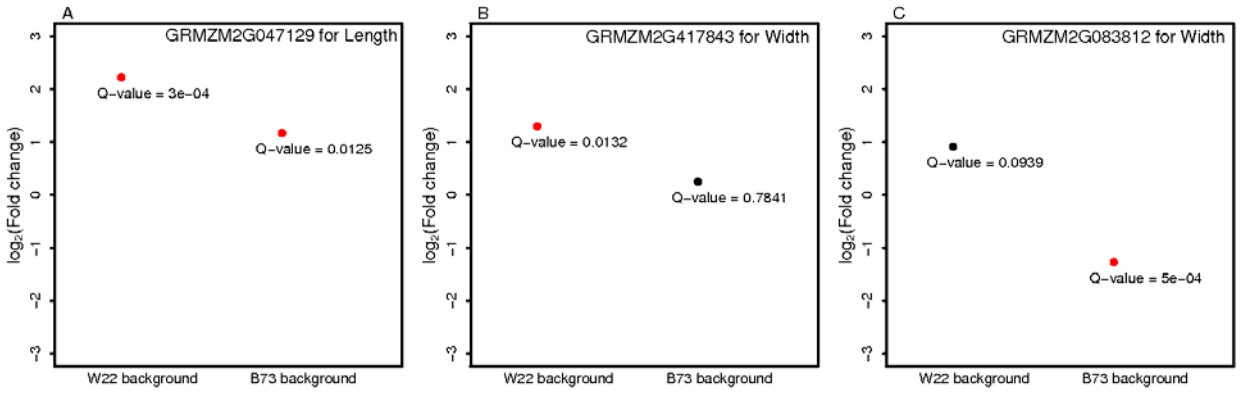
Supplementary Figure 5. The consistency of the genic and non-genic TAS distributions supported by simulations. In each simulation, a set of discovered TASs were randomly dropped out of the merged SNP data and the next highly associated SNPs within each target region were tabulated as the TASs. Even with missing over one-third of the TASs, the estimates of the percentage of non-genic TASs are still consistent with the findings from the whole dataset (red line). 1000 simulations were conducted for each case. Percentage of the non-genic SNP in the whole dataset is shown with the blue line.



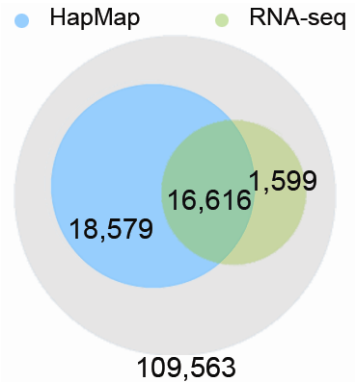
Supplementary Figure 6. Expression profiles of TAS-implicated candidate genes among developmental stages. The 10 stages along x-axis are four embryo (proembryo, transition phase, coleoptile stage, and L1 stage embryo; grey column 1 to 4); two meristems (SAM from two-week seedlings, and lateral meristem from two-week seedlings; white column 5 to 6), and four leaf gradient zones (basal zone, transitional zone, maturing zone, and mature zone; black column 7 to 10). Data for four leaf gradient zones are from Li et al. 2010.



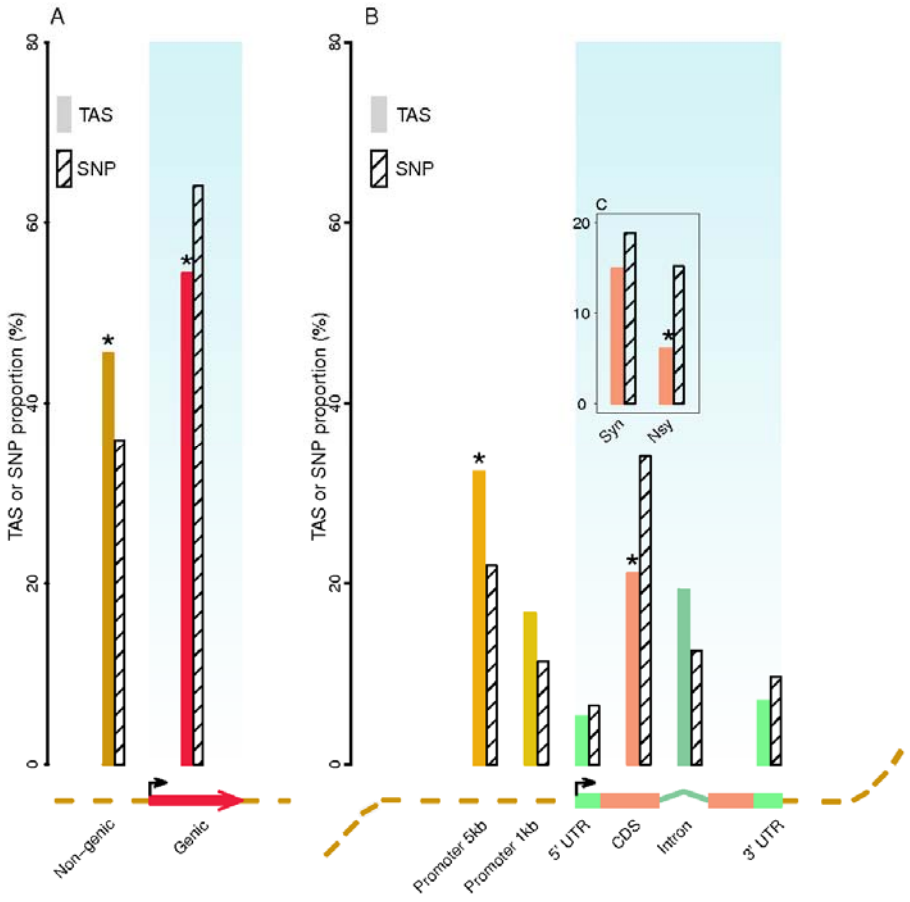
Supplementary Figure 7. Expression profile of two AP2 candidate genes. The expression of AP2 (GRMZM2G12977) associated with leaf length is higher than that of another AP2 (GRMZM2G421033) with flowering time among SAM and leaf developmental stages. Data for four leaf gradient zones are from Li et al. 2010.



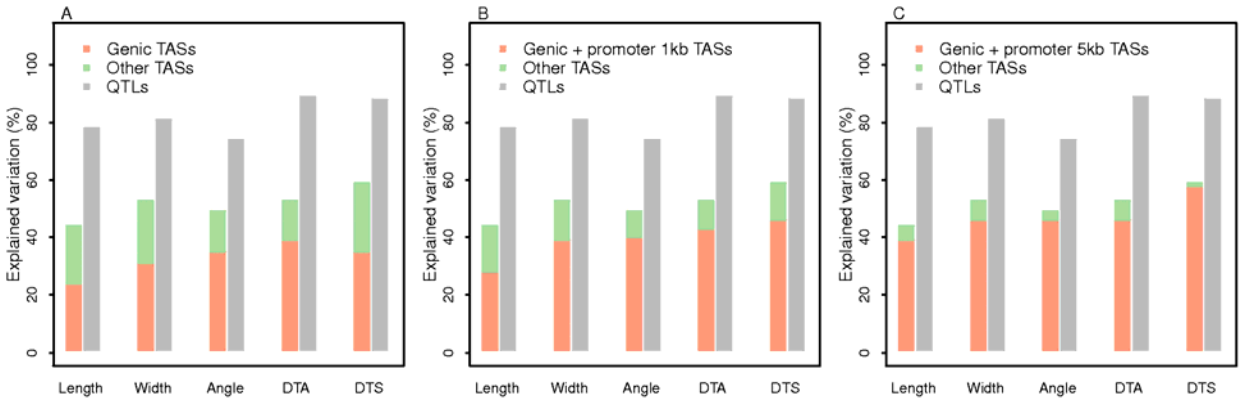
Supplementary Figure 8. Expression changes of three TAS-implicated genes between *lbl*1 and wide-type allele under W22 or B73 background. The red dot indicates the significant change and *Q*-value (adjusted *P*-value) is shown.



Supplementary Figure 9. Comparison between HapMap and RNA-seq of captured genes. Results are based on the maize Working Gene Set.



Supplementary Figure 10. Distribution of TASs and tested SNPs for five quantitative traits across genomic annotation sets. Results are based on the maize Working Gene Set. (A) non-genic versus genic; (B) different annotation sets; and (C) synonymous *versu*s nonsynonymous. Non-genic and promoter 5kb regions are over-represented among TASs. The nonsynonymous set is under-represented among TASs. Note that non-genic region includes promoter 5kb, which in turn includes promoter 1kb, and genic region includes the untranslated region (UTR), coding region (CDS), and intron. Stars denote the proportion of TASs from the annotation set differs significantly from that of tested SNPs. Arrow denotes transcription start site.



.

Supplementary Figure 11. Phenotypic variation explained by genic and non-genic TASs. Results are based on the maize Working Gene Set. (A) genic TASs versus all other TASs; (B) genic region plus promoter 1kb versus all other TASs, and (C) genic region plus promoter 5kb versus all other TASs. Phenotypic variation explained by all QTLs is shown for comparison.

Supplementary Table 1. The genomic location, statistical significance, allele frequency, and implicated genes of the identified TASs underlying five maize quantitative traits from the merged SNP dataset.

Trait	Ch	Position	-log$_{10}P$	-log$_{10}Q$	R^2	Freq[a]	Effect[b]	Gene	Annotation
Length	1	7,353,770	8.7	5.6	0.0087	0.04	12.12	GRMZM2G059428	No apical meristem protein
	1	15,122,404	4.6	1.5	0.0046	0.02	-16.17	-	-
	1	30,809,304	5.6	2.2	0.0053	0.16	-6.62	GRMZM2G072292	Expressed protein
	1	52,242,177	12.1	8.8	0.0133	0.29	6.69	GRMZM2G174896	Protein kinase
	1	85,130,587	20.7	17.9	0.0233	0.36	7.39	-	-
	1	244,100,533	23.7	21	0.0269	0.11	-13.68	GRMZM2G145458	ZIM domain
	1	283,982,913	7.6	4.9	0.0088	0.06	9.74	GRMZM2G361220	Clathrin adaptor complex small chain domain
	2	2,618,178	8.9	6.1	0.0106	0.04	-13.09	GRMZM2G065012	Expressed protein
	2	13,245,188	16.8	13.7	0.0185	0.37	-6.12	-	-
	2	59,584,754	21	17.4	0.0239	0.07	14.36	GRMZM2G134235	Expressed protein
	2	210,128,410	10	7.4	0.0103	0.41	4.35	-	-
	2	229,511,399	8.1	4.6	0.0074	0.30	-4.33	GRMZM2G444533	F-box and FBD domain
	3	8,019,525	7.7	5.5	0.0082	0.02	22.64	-	-
	3	22,034,011	10.9	7.1	0.0113	0.19	-7.87	GRMZM2G129777	AP2 domain
	3	177,565,419	5.8	2.3	0.0055	0.16	5.01	-	-
	3	212,248,700	46.8	43.4	0.0496	0.42	8.96	GRMZM2G047129	GDSL-like lipase
	4	3,248,688	21.7	18.5	0.0253	0.38	6.58	GRMZM2G097981	Tropinone reductase
	4	151,524,406	9.6	6.3	0.0096	0.06	-10.11	GRMZM2G323719	Amino acid kinase
	4	197,993,592	10.5	7.3	0.0114	0.05	-10.82	-	-
	5	3,062,680	12	9.1	0.0132	0.28	-5.50	GRMZM2G096165	Ribosomal protein
	5	74,510,491	17.8	14.8	0.0192	0.16	-11.19	-	-
	5	202,073,473	23.9	21.1	0.0261	0.45	-6.00	-	-
	6	23,864,967	7.4	5	0.0076	0.02	-15.71	GRMZM2G459581	MAP kinase phosphatase
	6	105,119,528	7.4	4	0.0074	0.07	8.85	GRMZM2G331283	Amino acid transporter
	6	152,684,138	8	4.3	0.0085	0.28	-4.46	GRMZM2G069737	F-box domain
	7	17,118,117	6.8	2.7	0.0073	0.10	-7.25	-	-
	7	150,990,249	6.5	3.6	0.0071	0.11	-7.05	-	-
	7	161,864,686	10.5	6.5	0.0122	0.26	5.69	-	-
	8	5,776,641	7.5	4.5	0.0073	0.04	-10.98	GRMZM2G155260	Protein kinase
	8	121,054,240	18.2	15.1	0.0209	0.14	-10.86	-	-
	8	164,043,117	6.7	3.6	0.0074	0.08	8.08	-	-
	9	13,032,982	10.5	7.5	0.0112	0.22	5.89	-	-
	9	104,339,779	12	8.6	0.0127	0.15	9.09	-	-
	9	131,683,423	24.5	21.8	0.0288	0.31	9.17	-	-
	10	13,610,740	6.7	4.2	0.0056	0.08	-6.73	GRMZM2G125931	Expressed protein
	10	132,894,369	13.5	10.4	0.0145	0.12	-10.71	GRMZM2G425072	Auxin-responsive SAUR gene
	10	148,546,034	9.8	6.6	0.0108	0.21	-5.31	GRMZM2G039011	Expressed protein
Width	1	21,725,298	32.4	29.7	0.0125	0.19	1.33	-	-
	1	23,250,177	30.4	27.2	0.0016	0.36	0.40	-	-
	1	55,604,358	10.6	7.8	0.0112	0.29	0.85	GRMZM2G075262	Expressed protein
	1	160,331,991	6.9	3.6	0.0070	0.04	2.05	-	-
	1	210,399,194	5.1	2.1	0.0055	0.04	-1.44	-	-
	1	277,518,310	16.2	12.9	0.0172	0.18	1.45	GRMZM2G330218	Expressed protein
	1	298,083,583	7	3.9	0.0078	0.17	-0.81	GRMZM2G417843	4-nitrophenylphosphatas
	2	10,679,395	7.7	4.8	0.0080	0.10	-0.98	-	-

	2	57,602,831	106.1	104.1	0.1153	0.44	1.90	AC191050.3_FG003	Globulin-1 S allele
	2	124,883,226	106.7	104.3	0.1161	0.42	1.95	-	-
	2	214,076,426	12.3	9.4	0.0131	0.37	0.64	GRMZM2G030646	Eukaryotic translation initiation factor 2
	3	5,633,931	6.7	3.3	0.0077	0.04	-1.68	-	-
	3	31,698,831	34.3	30.6	0.0376	0.39	1.26	GRMZM2G049672	Zinc finger, C3HC4 type
	3	163,889,745	24.8	21.7	0.0273	0.29	-1.30	-	-
	3	194,352,239	10.2	7.4	0.0107	0.21	-0.90	-	-
	3	204,956,307	11.5	7.7	0.0013	0.07	-0.62	-	-
	4	156,002,632	12.9	9.1	0.0134	0.11	-1.31	-	-
	4	201,525,973	11.2	9	0.0125	0.13	-1.06	GRMZM2G163709	Phospholipase C
	5	6,130,146	26.7	23.6	0.0291	0.36	-1.01	GRMZM2G171468	MYB transcription factor
	5	31,528,232	69.6	64.5	0.0776	0.48	-1.56	GRMZM2G337048	UDP-glucosyl transferase
	5	185,911,974	5.1	1.6	0.0055	0.08	-0.97	-	-
	5	213,362,996	12	8.6	0.0139	0.09	1.35	GRMZM2G083812	Heparanase-like
	6	119,603,988	25.7	23.6	0.0283	0.27	-1.39	GRMZM2G075710	Embryogenesis transmembrane protein
	6	147,218,755	19.1	16.3	0.0209	0.34	-1.17	-	-
	6	157,675,994	7.2	3.6	0.0069	0.11	1.08	-	-
	6	165,339,108	9	6.2	0.0088	0.35	-0.62	-	-
	7	3,576,083	13.8	10	0.0159	0.06	1.72	-	-
	7	121,502,283	10.7	6.8	0.0109	0.25	0.74	GRMZM2G072218	Serine Carboxypeptidase
	7	165,920,367	11.5	8.3	0.0133	0.18	0.88	GRMZM2G137064	PPR repeat domain
	8	22,860,690	9.8	6.6	0.0100	0.21	-0.88	-	-
	8	121,052,910	13	10.1	0.0147	0.17	1.24	-	-
	8	167,519,904	7.9	4.4	0.0083	0.17	-0.78	-	-
	9	121,382,512	32.8	30.2	0.0359	0.28	1.31	GRMZM2G131221	Expressed protein
	10	124,313,332	11.9	10	0.0124	0.02	2.65	-	-
	10	124,863,499	12.1	10.2	0.0125	0.02	2.66	-	-
	10	147,266,926	18.9	15.7	0.0214	0.40	-0.80	GRMZM2G475882	Auxin response factor
Angle	1	30,342,426	29.7	26.7	0.0315	0.47	-0.83	-	-
	1	251,379,569	20.1	16.5	0.0208	0.38	0.82	GRMZM2G014392	9-cis-epoxycarotenoid dioxygenase
	1	281,810,922	6.4	3.1	0.0064	0.18	-0.62	GRMZM2G068158	Methyltransferase
	2	4,156,772	42.9	40.1	0.0452	0.42	-1.11	GRMZM2G039532	Expressed protein
	2	14,696,322	13.5	10.2	0.0130	0.31	0.68	-	-
	2	49,478,060	7.1	4.1	0.0067	0.02	-1.73	-	-
	2	209,607,885	5.2	2.2	0.0049	0.18	0.49	GRMZM2G064630	MYB transcription factor
	3	31,703,959	20.6	17	0.0213	0.42	0.82	-	-
	3	158,616,899	22.9	20.6	0.0237	0.06	-2.49	-	-
	3	175,048,356	51.8	49.4	0.0557	0.36	-1.60	GRMZM2G060216	Liguleless 2
	3	202,365,265	5.2	1.9	0.0040	0.10	0.62	-	-
	4	129,385,200	21.3	18.2	0.0217	0.13	-1.30	GRMZM2G104118	Glycosyl transferase
	4	180,859,297	5.3	1.9	0.0055	0.04	1.20	-	-
	4	242,646,294	8.5	5.4	0.0082	0.08	1.00	GRMZM2G012143	SETH3-like
	5	3,193,230	13.1	9.9	0.0137	0.23	-0.75	-	-
	5	33,281,655	61.1	59.4	0.0643	0.15	-2.22	GRMZM2G075150	Exocyst complex component
	5	74,060,756	48.9	46.5	0.0527	0.31	-1.40	-	-
	5	199,130,596	20	17.4	0.0209	0.40	-0.70	-	-
	6	16,432,104	5.8	2.6	0.0053	0.11	0.81	-	-
	6	91,074,658	25.4	22.2	0.0133	0.06	2.02	-	-

	7	128,396,999	11.5	9.1	0.0115	0.13	-0.89	-	-
	8	112,526,887	29.5	27	0.0306	0.02	-3.87	GRMZM2G139574	Glycosyltransferase
	8	117,449,528	27.4	25.2	0.0283	0.02	-3.80	GRMZM2G113554	Expressed protein
	8	166,486,198	13.8	11	0.0142	0.36	-0.62	-	-
	9	105,203,693	23.3	20.8	0.0253	0.02	-3.68	-	-
	9	142,011,299	21.4	17.5	0.0227	0.29	-0.91	GRMZM2G045981	Leucine-rich repeat protein
	10	116,704,001	6.3	3.3	0.0056	0.07	-0.87	-	-
	10	143,270,554	7.6	4.8	0.0080	0.33	-0.49	GRMZM2G180471	Phosphatase 2C
DTA	1	15,859,601	11.2	7.9	0.0129	0.11	0.38	GRMZM2G000741	Mitochondrial carrier
	1	45,309,466	14	11	0.0157	0.40	0.23	-	-
	1	198,503,585	8.4	5.2	0.0080	0.35	-0.19	GRMZM2G001803	Expressed protein
	1	223,148,600	6.6	3.8	0.0071	0.14	0.26	-	-
	1	285,088,949	10.1	7	0.0114	0.19	-0.25	GRMZM2G174136	Expressed protein
	2	13,183,762	12.6	9.4	0.0139	0.22	-0.27	GRMZM2G035719	RNA recognition motif
	2	71,694,864	23.3	19.8	0.0252	0.04	0.79	GRMZM2G000980	Putative uncharacterized protein
	2	196,646,460	7.3	4.1	0.0074	0.13	0.29	-	-
	2	215,685,793	34.1	31.4	0.0347	0.29	0.39	-	-
	3	33,704,604	30.7	27.8	0.0323	0.36	0.37	-	-
	3	109,874,104	32.9	31.1	0.0373	0.28	0.49	-	-
	3	157,651,234	19.6	16.5	0.0237	0.23	-0.47	GRMZM2G171600	Calmodulin-binding transcription activator
	3	197,449,919	21.8	19.3	0.0246	0.31	0.34	GRMZM2G177906	Hexokinase
	3	213,247,367	25.3	22.1	0.0253	0.41	0.30	GRMZM2G067583	Tat pathway signal sequence family
	4	39,171,288	10.8	8.8	0.0097	0.08	0.36	GRMZM2G005583	Esterase
	4	218,097,053	9.4	6.2	0.0103	0.03	0.54	-	-
	5	3,379,634	13.4	10.6	0.0127	0.33	-0.21	-	-
	5	129,802,132	17.2	14.1	0.0214	0.19	0.35	-	-
	6	58,682,095	7.8	5.1	0.0071	0.04	-0.41	-	-
	6	155,742,011	5.9	2	0.0051	0.11	0.25	GRMZM2G116258	Glutamate-1-semialdehyde-2,1-aminomutase
	6	162,708,977	17.6	14.7	0.0192	0.17	0.38	GRMZM2G061663	Zinc finger transcription factor
	7	145,739,720	10.9	7.7	0.0135	0.14	-0.33	-	-
	7	168,992,525	12.3	9	0.0136	0.36	0.20	GRMZM2G134341	PRM5
	8	121,051,703	36.1	33.6	0.0399	0.25	0.48	GRMZM2G527250	Expressed protein
	8	160,241,887	10.3	7.8	0.0107	0.25	0.25	-	-
	9	26,551,727	5.2	2.3	0.0051	0.04	0.48	-	-
	9	113,530,209	62.3	60	0.0320	0.31	0.45	-	-
	9	124,069,745	39.6	37.5	0.0404	0.47	0.40	-	-
	9	142,535,940	6.3	3.6	0.0057	0.05	-0.38	-	-
	10	8,398,599	6.4	3	0.0064	0.20	0.21	GRMZM2G136910	Abscisic stress-ripening
	10	87,615,338	42.5	39.6	0.0452	0.07	0.92	GRMZM2G062541	Helix-loop-helix DNA-binding domain
	10	144,082,871	11.6	8.8	0.0121	0.31	-0.21	GRMZM2G445575	Transcription factor
DTS	1	21,691,747	8.7	5.8	0.0077	0.22	0.25	GRMZM2G016922	Terpene synthase
	1	42,970,480	11.5	8.4	0.0124	0.10	0.52	-	-
	1	196,441,970	5.8	2.5	0.0048	0.15	-0.24	GRMZM2G045135	Haloacid dehalogenase-like hydrolase
	1	236,344,417	8.7	5.1	0.0087	0.22	0.26	GRMZM2G141256	3-oxoacyl-reductase
	1	286,872,569	10.7	7.7	0.0127	0.08	0.47	-	-
	2	13,183,762	20	16.5	0.0210	0.22	-0.38	GRMZM2G035719	RNA recognition motif

2	56,512,014	26.4	23.4	0.0284	0.29	0.38	GRMZM2G434669	Expressed protein
2	205,195,986	6.1	3.2	0.0058	0.08	0.44	GRMZM2G403800	Expressed protein
2	215,685,793	25.1	21.8	0.0268	0.29	0.41	-	-
3	23,542,690	17.1	13.8	0.0003	0.30	0.07	GRMZM2G114552	BBTI1
3	158,221,870	15.2	11.7	0.0188	0.12	-0.56	-	-
3	211,703,629	58.2	56.5	0.0142	0.48	0.27	-	-
3	212,725,952	58.2	56.4	0.0142	0.48	0.27	-	-
4	17,307,125	12.8	9.6	0.0123	0.04	0.78	-	-
4	155,219,577	8.1	5	0.0082	0.09	-0.42	GRMZM2G003514	Zea agamous5
4	221,605,444	10	8	0.0095	0.01	0.85	-	-
5	2,583,875	12.8	9.7	0.0133	0.31	-0.25	-	-
5	159,438,987	11	7.7	0.0114	0.09	0.41	-	-
5	207,746,609	9.2	6.2	0.0086	0.08	0.40	-	-
6	103,280,254	8.6	6.2	0.0063	0.04	0.50	-	-
6	160,828,364	14.1	11	0.0146	0.11	0.46	GRMZM2G474546	Protein kinase
6	166,267,781	7.9	4.8	0.0078	0.31	0.21	GRMZM2G114578	AMP-binding enzyme
7	138,927,147	13	9.8	0.0140	0.33	-0.25	-	-
7	168,992,525	7.3	4.3	0.0085	0.36	0.18	GRMZM2G134341	PRM5
8	6,989,621	9.7	7.6	0.0087	0.48	-0.16	GRMZM2G109627	No apical meristem protein
8	161,249,047	16.5	13.4	0.0185	0.24	0.36	-	-
9	93,250,059	8.7	5.1	0.0097	0.06	0.60	GRMZM2G141256	3-oxoacyl-reductase
9	124,365,160	43.9	41.4	0.0473	0.31	0.61	GRMZM2G095124	Adapitin protein
9	142,281,937	7.5	4.2	0.0082	0.14	-0.31	GRMZM2G102218	YABBY domain
10	87,615,338	58.7	55.9	0.0657	0.07	1.13	GRMZM2G062541	Helix-loop-helix DNA-binding domain
10	144,836,415	10.9	7.9	0.0132	0.28	-0.25	GRMZM2G421033	AP2 domain

[a] Frequency of non-B73 allele within the maize NAM population. The range of the theoretical frequency for non-B73 allele is 2% - 50% with the NAM design. [b] Genetic effect as compared with B73 allele. Leaf length, mm; leaf width, mm; leaf angle, degree; DTA, days to anthesis, day; and DTS, days to silking, day.