# Joint linkage–linkage disequilibrium mapping is a powerful approach to detecting quantitative trait loci underlying drought tolerance in maize

Yanli Lu[a,b], Shihuang Zhang[c], Trushar Shah[b,1], Chuanxiao Xie[c], Zhuanfang Hao[c], Xinhai Li[c], Mohammad Farkhari[b,d], Jean-Marcel Ribaut[e], Moju Cao[a], Tingzhao Rong[a], and Yunbi Xu[b,f,2]

[a]Maize Research Institute, Sichuan Agricultural University, Ya'an 625014, China; [b]International Maize and Wheat Improvement Center, El Batán, 56130, Mexico; [c]Institute of Crop Science, National Key Facilities for Crop Genetic Resources and Improvement, Chinese Academy of Agricultural Sciences, Beijing 100081, China; [d]Plant Breeding Department, College of Agriculture, University of Tehran, Karaj 31587-77871, Iran; [e]Generation Challenge Programme c/o CIMMYT, El Batán, 56130, Mexico; and [f]Institute of Crop Science/International Maize and Wheat Improvement Center, National Key Facilities for Crop Genetic Resources and Improvement, Chinese Academy of Agricultural Sciences, Beijing 100081, China

This paper describes two joint linkage–linkage disequilibrium (LD) mapping approaches: parallel mapping (independent linkage and LD analysis) and integrated mapping (datasets analyzed in combination). These approaches were achieved using 2,052 single nucleotide polymorphism (SNP) markers, including 659 SNPs developed from drought-response candidate genes, screened across three recombinant inbred line (RIL) populations and 305 diverse inbred lines, with anthesis-silking interval (ASI), an important trait for maize drought tolerance, as the target trait. Mapping efficiency was improved significantly due to increased population size and allele diversity and balanced allele frequencies. Integrated mapping identified 18 additional quantitative trait loci (QTL) not detected by parallel mapping. The use of haplotypes improved mapping efficiency, with the sum of phenotypic variation explained (PVE) increasing from 5.4% to 23.3% for single SNP-based analysis. Integrated mapping with haplotype further improved the mapping efficiency, and the most significant QTL had a PVE of up to 34.7%. Normal allele frequencies for 113 of 277 (40.8%) SNPs with minor allele frequency (<5%) in 305 lines were recovered in three RIL populations, three of which were significantly associated with ASI. The candidate genes identified by two significant haplotype loci included one for a SET domain protein involved in the control of flowering time and the other encoding aldo/keto reductase associated with detoxification pathways that contribute to cellular damage due to environmental stress. Joint linkage–LD mapping is a powerful approach for detecting QTL underlying complex traits, including drought tolerance.

anthesis-silking interval | drought resistance | haplotype loci | integrated quantitative trait locus mapping

Quantitative trait loci (QTL) for complex traits in plants generally have been mapped through biparental population–based linkage analysis or diverse germplasm–based linkage disequilibrium (LD) analysis (1, 2). LD studies have significant potential to identify the genes responsible for a particular phenotype, and, unlike biparental populations, have the power to simultaneously evaluate the varying effects of many alleles (3). In maize, due to the rather low level of LD, only polymorphisms separated from a locus responsible for the phenotypic effect by a few hundred bases are likely to be significantly associated with variation for a trait in a randomly mated population. On one hand, a linkage approach is powerful for detecting genetic effects at loci involved in the expression of target traits. On the other hand, LD mapping offers the ability to exploit all recombination events that have occurred in the evolutionary history of a sample set of germplasm, allowing for increased mapping resolution.

One of the key limiting factors of LD mapping is the effect of genetic relatedness in the population under study, which often causes the identification of spurious associations (4, 5). Excluding these effects from the LD analysis (6, 7) can be commonly achieved using the STRUCTURE program (8, 9), principal component analysis (10), or random molecular markers to estimate pairwise relatedness among all individuals via a mixed linear model (5). The mixed linear model method has been successfully applied in LD mapping of many crops, including maize (5, 11, 12). Another limitation of LD mapping is related to rare alleles that severely limit the power of QTL detection. Low-frequency functional alleles are among the likely culprits for the major missing components of the heritable variation in humans (13). Given that most alleles are rare in a large proportion of accessions from diverse germplasm collections, it would be difficult to account for most of the phenotypic variation in a target trait using LD mapping.

Joint linkage–LD mapping is an alternative approach to overcoming some of the inherent limitations of both linkage and LD methods (1). Several statistical techniques have been developed for this purpose (14–16). Following that rationale, the maize community has developed the nested association mapping population (17), in which the use of controlled crosses reduces the confounding effects of population structure, whereas the large numbers of progeny derived from the 25 crosses allows for linkage mapping with substantial statistical power. The number of founder lines required to capture the desired amount of genetic variation remains unclear, however. The use of a large number of inbred lines combined with several biparental populations could be considered as an alternative. We explore this alternative in the present work.

Drought at any stage of plant development affects grain production, but causes maximum damage in maize when it occurs around the time of flowering (18). A commonly used secondary trait for drought tolerance in maize is the asynchrony between silk emergence and pollen shedding. Under water-limited conditions, this asynchrony, termed the anthesis-silking interval (ASI), is highly correlated with grain yield (19), which has been extensively studied through linkage analysis (20).

The present study was conducted to investigate opportunities for resolving some limiting factors for linkage and LD mapping

PLANT BIOLOGY

through joint linkage–LD mapping and candidate gene-based approach using three RIL populations and one inbred line population. The target trait ASI was scored under both well-watered (WW) and water-stressed (WS) conditions. The results indicate that joint linkage–LD mapping combined with the use of candidate genes is a powerful approach to detecting QTL underlying complex traits.

## Results

**Allele Diversity and Population Structure.** From two SNP chips, a total of 2,052 high-quality SNPs were selected using 522 maize lines included in this study (305 inbred lines and 217 RILs from three populations). There was an even distribution of minor allele frequency across the 2,052 SNPs (Fig. 1A), and 11.3% (232) of SNP markers had a minor allele frequency of >5% across all tested lines, with a slightly higher percentage in the inbred line set ($n = 305$; Fig. 1B). Using all SNPs within each 10-kb block, 386 haplotype loci were identified, each containing 2–12 SNPs and 2–23 alleles per haplotype locus (Table S1). Dataset S1 and Table S2 provide distributions of haplotypes on the genome and summary information for each chromosome. Polymorphic information content for the 386 haplotypic loci varied dramatically, from 0.015 to 0.828 (average, 0.434). Five groups within the entire line set were identified by principal component analysis, with well-separated temperate and tropical maize lines (Fig. S1A). For the inbred line set ($n = 305$), three genetically distinct groups were identified (Fig. S1B). In conclusion, structure analysis separated clearly biparent-derived populations from the inbred population, and thus the structure-related confounding effect on marker–trait association analysis should be largely removed.
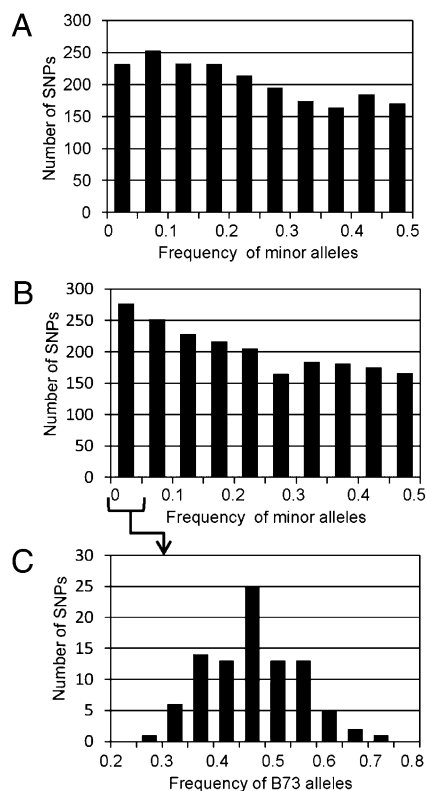


**Fig. 1.** Distribution of minor allele frequencies for 522 lines (A) and 305 lines (B) at 2,052 SNP loci and distribution of allele frequency for B73 alleles at the 93 SNP loci with minor allele frequency <0.05 in 305 maize lines (C).

**QTL Identified by Linkage Mapping.** Of the 2,052 SNPs selected, 1,991 were mapped in silico to the maize physical map and used for QTL linkage mapping using phenotypic extremes from each of the three RIL populations, C5, C6, and XB. The total length of the linkage maps was 1,652 cM for C5, 1,629 cM for C6, and 1,668 cM for XB, with respective average distances between markers of 3.89, 3.44, and 2.34 cM.

Both single-marker analysis and composite interval mapping (CIM) were used for QTL mapping with each RIL population. Single-marker analysis identified one QTL for ASI-WW condition and four QTL for the ASI-WS condition, including 17 single SNP markers (Table S3). Eight QTL for ASI were identified across the two water regimes using CIM (Table S4). One, four, and three QTL were identified using the C5, C6, and XB populations, respectively (Fig. S2). Parental lines Ac7643, CML444, and B73 had alleles with a negative effect (i.e., reducing ASI) at one, three, and two QTL, respectively. Conversely, both CML444 and B73 had positive alleles (i.e., increasing ASI) at one QTL.

Some QTL for different water regimes were identified in the same or neighboring chromosome regions (Fig. S3). Single-marker analysis showed good agreement with CIM, with four QTL (one on chromosome 2, two on chromosome 4, and one on chromosome 9) identified by both methods. For the other four QTL identified by CIM, associated markers had relatively high LOD scores in single-marker analysis. We note that the relatively small population sizes used in linkage mapping via selective genotyping might have underestimated QTL numbers but overestimated the sum of phenotypic variation explained (PVE).

**QTL Identified by LD Mapping Using Single SNP and Haplotype Data.** LD mapping was performed with a mixed linear model by integrating population structure and family relatedness within populations using both 2,052 SNPs and 386 haplotype loci with rare alleles excluded. To correct for multiple comparisons, a Bonferroni-corrected threshold probability based on individual tests (21) was calculated. Using single SNP–based LD mapping, no significant marker was identified at the Bonferroni-corrected threshold ($-\log P > 5.31$ for 2,052 SNPs), and thus associations are presented at the suggestive level of significance ($P < 0.001$; $-\log P > 3.00$). Using the 305-line set, three markers were identified for WW and one marker was identified for WS, whereas using the 522-line set, four markers were identified for WW and three markers were identified for WS (Tables S5 and S6). Only one marker for WW (PZA03109.1) was shared between the 522- and 305-line sets.

Using the 305-line set, haplotype-based LD mapping identified two haplotype loci (HP153 and HP312) associated with ASI-WW at $P < 0.001$, with 3.66% and 4.21% of PVE, respectively (Table 1 and Table S5). Using the 522-line set, three haplotype loci (two for ASI-WW and one for ASI-WS) were identified. Relatively high $-\log P$ and $R^2$ values were obtained for 305-line set at the locations where a suggested significance was detected for the 522-line set, or vice versa (Table 1). Each of two large-effect haplotypes, HP71 for ASI-WW on chromosome 2 and HP322 for ASI-WS on chromosome 8, had >20% of PVE (Table 1) and were significant at the Bonferroni-corrected level. The locus HP322 included two closely linked SNPs, neither of which was significant in terms of linkage or LD mapping; however, one of the markers (PZA00142.4) had a relatively high $-\log P$ value (1.30). Another haplotype locus (HP71), including 10 SNPs from two genes, detected 11 alleles in the 522-line set, with 22.70% of PVE for ASI-WW. Significant differences in ASI (3.07–6.25) were found among these 11 allelic classes (Table 2).

In conclusion, haplotypes combining two or more SNPs had a higher PVE, whereas each single SNP had a lower PVE. Considering all markers that were significant at either the suggested or Bonferroni-corrected level, the average PVE increased

**Table 1. Haplotype loci significantly associated with ASI under WW and WS regimes as revealed by LD mapping using 522 and 305 maize lines and integrated linkage–LD mapping using 522 lines**

| Trait | Locus | Chr | Bin | No. of SNPs | No. of alleles | LD mapping Population | LD mapping −Log $P$ | LD mapping $R^2$, % | Integrated mapping −Log $P$ | Integrated mapping $R^2$, % | SNPs included |
|---|---|---|---|---|---|---|---|---|---|---|---|
| ASI-WW | HP71 | 2 | 2.03 | 10 | 11 | 522, 305 | 8.71, 1.66 | 22.70, 7.30 | 12.12 | 34.69 | PZA03747.7, PZA03747.4, PZA03747.2, PZA03747.1, PZA03748.2, PZA03748.3, PZA03748.1, PZA03701.1, PZA03700.2, PZA03699.1 |
| | HP153 | 4 | 4.02 | 2 | 4 | 522, 305 | 3.18, 4.14 | 2.30, 3.66 | 3.18 | 2.30 | PZA00136.2, PZA03109.1 |
| | HP312 | 8 | 8.06 | 2 | 3 | 522, 305 | 1.79, 3.64 | 1.28, 4.21 | 2.85 | 2.04 | PZB01385.3, PZB01385.2 |
| ASI-WS | HP322 | 8 | 8.06 | 2 | 3 | 522, 305 | 8.30, 1.16 | 21.50, 4.42 | 22.06 | 28.81 | PZA00142.3, PZA00142.4 |

$R^2$, explained phenotypic variance. The threshold for multiple tests is −log $P$ > 4.59, for 386 haplotype loci.

from 1.54% for single SNP–based LD mapping to 17.00% for haplotype-based LD mapping with the 522-line set, whereas the sum of PVE increased from 5.38% to 23.25% and −log $P$ increased from 3.24 to 7.12 (Table S5). Improved LD mapping efficiency using haplotypes is shown by quantile–quantile (Q–Q) plots (Fig. 2), in which observed and expected probability distributions are compared by plotting their quantiles against each other (22) and a marker–trait association is indicated by the deviation of SNP–trait data points from the null hypothesis (1:1 diagonal line). For single SNP–based analysis, no association with a $P$ value deviated noticeably from the null hypothesis to reach the Bonferroni-corrected threshold (−log $P$ > 5.31 for 2,052 SNPs), although seven data points (four for ASI-WW and three for ASI-WS) were significant at the suggested level (−log $P$ > 3.00) (Fig. 2A). For haplotype-based analysis, however, HP71-ASI-WW and HP322-ASI-WS had $P$ values that deviated significantly from the diagonal line and were much higher than the horizontal line (−log $P$ = 4.59 for 386 haplotype loci) (Fig. 2B). The plots in Fig. 2B also show that the mixed model was quite effective in accounting for population structure and relative kinship, as indicated by the majority of marker–trait data points close to the 1:1 line for $P_{observed}$ versus $P_{expected}$.

**QTL Identified by Integrated Linkage–LD Mapping.** Integrated mapping, in which datasets from the three RIL populations and one inbred-line population were merged for a single analysis, was implemented to identify the best of the eight genetic models (all

**Table 2. Phenotypic differences for ASI-WW contributed by the haplotye locus HP71 in haplotype-based LD mapping using 522 maize lines**

| | Haplotype* | $n$ | ASI, mean ± SD |
|---|---|---|---|
| 1 | AAG**AA**A**GGG**A | 15 | 3.07 ± 3.15 |
| 2 | **G**AGAGA**GGG**C | 8 | 3.14 ± 3.35 |
| 3 | AA**AA**G**GGGG**A | 6 | 3.92 ± 2.31 |
| 4 | AA**AA**G**GG**A**G**A | 12 | 4.22 ± 3.89 |
| 5 | A**TG**A**GGG**A**GC** | 27 | 4.33 ± 4.79 |
| 6 | A**T**GA**GA**G**GG**A | 9 | 4.50 ± 1.81 |
| 7 | AA**G**GGA**GA**G**C** | 57 | 4.65 ± 4.89 |
| 8 | AA**G**GGAAA**G**A | 85 | 5.16 ± 3.64 |
| 9 | **G**AGAGA**G**A**G**A | 30 | 5.33 ± 4.42 |
| 10 | A**TG**A**G**A**G**A**G**A | 4 | 5.88 ± 5.11 |
| 11 | AAG**A**GA**G**AAA | 6 | 6.25 ± 1.70 |

*From left to right, each character represents a nucleotide from a SNP marker: PZA03747.7, PZA03747.4, PZA03747.2, PZA03747.1, PZA03748.2, PZA03748.3, PZA03748.1, PZA03701.1, PZA03700.2, and PZA03699.1. The explained phenotypic variance for ASI-WW is 22.70%. Nucleotides in bold type and underscored have a negative effect on ASI.

possible combinations of QTL allele effects in these four populations) using both single markers and haplotypes. For single SNP–based integrated mapping, no significant association was found for any model at the Bonferroni-corrected threshold (Table S6). At the suggested level, 21 markers were identified, 18 of which only by integrated mapping.

The identification of 18 additional potentially significant markers by integrated mapping can be attributed mainly to genetic effects, background effects, or both, contributed by the RIL populations (Dataset S2A and Table S6). For example, of the 10 of 21 markers with an apparent genetic effect contributed by one RIL population ($R^2$ >1%), seven had the same allele effect as the inbred population, whereas three had different allele effects. The 18 additional markers also could be partially attributed to the genetic effect contributed by the 305-line set, as indicated by their relatively large −log $P$ and $R^2$ values. For the five markers, one from each QTL region identified by single marker–based linkage analysis (Tables S3 and S6), relatively high −log $P$ values were revealed by integrated mapping with specific genetic models. In four of the five cases, −log $P$ values (1.44–2.03) were increased significantly compared with the LD mapping with 305 lines (0.21–1.37) (Table S6).

For the 11 markers identified by integrated mapping for ASI-WW at the suggested level, 64 metahaplotype alleles (across all significant loci, including those with physical distances >10 kb or located in different chromosomes) were identified with 18.27% of PVE (−log $P$ = 3.10). Although each single SNP had only a small PVE (1.37–2.12%), the allele combination had a much higher PVE.

Haplotype-based integrated mapping confirmed the two haplotype loci revealed by LD mapping. For HP71, all of the RIL populations had opposite allele effects to those in the inbred set, resulting in an increase in $R^2$ from 22.7% to 34.7%. For HP322, the populations C5 and C6 had the opposite allele effect to those in the inbred set, resulting in an increase in $R^2$ from 21.5% to 28.8% (Table 1).

Several clusters of significant SNP markers were identified by both linkage and LD mapping (Fig. S3). For example, a region on chromosome 2 spanning 16.2 cM included two QTL for ASI-WS identified by CIM and LD mapping and one QTL for ASI-WW identified by LD mapping. The three chromosome maps also included five SNPs identified by integrated mapping, four of which were for ASI-WW (Fig. S3). Nine QTL were common across both water regimes and more than one QTL mapping approach.

A total of 277 SNPs were excluded from LD analysis due to minor allele frequency of <5% in the 305-line set (Fig. 1B). Of these 277 SNPs, 93 were polymorphic in the XB population with normal allele frequencies recovered (Fig. 1C). One of these
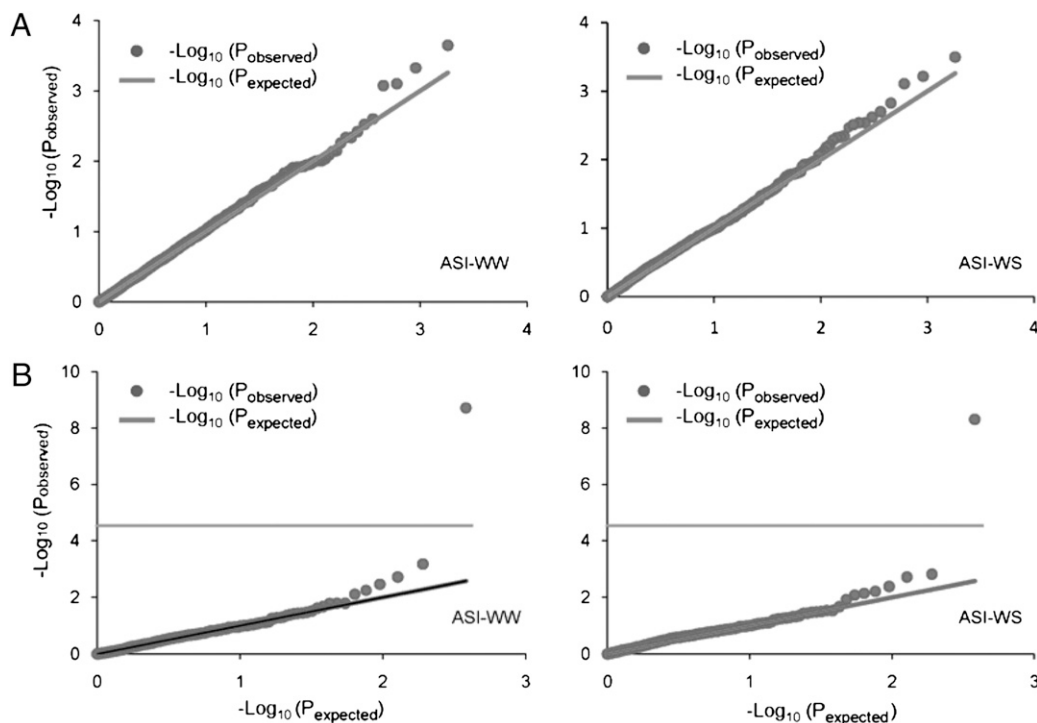
**Fig. 2.** Q–Q plots for SNP markers significantly associated with ASI under WW and WS conditions using 522 maize lines. (*A*) Single SNP–based analysis. (*B*) Haplotype-based analysis using 386 haplotype loci.

polymorphic markers (umc13.1 on chromosome 1) was found to be associated with ASI-WS by single marker–based linkage analysis, and two markers, PZA00344.10 and PAZ00545.26 (chromosomes 4 and 5), were associated with two QTL for ASI-WW and one QTL for ASI-WS by CIM.

**Candidate Genes for Drought Tolerance.** The candidate genes for ASI were identified using associated SNP markers detected by different mapping approaches (Dataset S24). The identified genes include those involved in cytochrome P450 (*CYP71C1, CYP71C2, CYP71C3*, and *CYP90A21*), methylation (a gene related to the methyl-CpG–binding domain, MBD101, and a SET [Su (var), Enhancer of zeste (E(z)), and Trithorax] domain protein (SDG140), transcription factors (*zmm19*, *myb42*, C2H2-type zinc finger protein), dehydrins (late embryogenesis-abundant protein and heat-shock proteins as molecular chaperones of dehydrins), osmotins (trehalose-6-phosphate synthase), oxidoreducases (aldo/keto reductase), and signal transduction factors (serine/threonine protein kinase and protein phosphatase 2C). These genes are frequently associated with abiotic stress, including drought (23).

One of the two most significant haplotype loci, HP71, contains seven linked SNPs derived from one gene (GRMZM2G164400) with unknown function and three SNPs from a second gene (SDG140). SDG140 is a SET domain protein in maize that is generally associated with methyltransferase activity targeted to specific lysine residues of histone H3 or H4. Regulation of histone methylation by SET domain–containing methyltransferases is completed by protein–protein interactions through both intramolecular and intermolecular associations that are important in plant developmental processes (24), including two major component traits of ASI.

The second significant haplotype locus, HP322, includes two closely linked SNPs from a gene encoding aldo/keto reductase (AKR). In addition, PZA02792.25, identified by integrated mapping with 522 lines and LD mapping with 305 lines, is also located in a gene of the AKR family. AKRs can detoxify lipid peroxidation products and glycolysis-derived reactive aldehydes that contribute significantly to cellular damage caused by environmental stress (25). The specific members of this NADPH-dependent AKR superfamily play important roles in the synthesis of sugar alcohols (e.g., sorbitol, mannitol). The products of these reactions can act as radical scavengers even at low concentrations, and their accumulation as osmolytes can lead to osmotic adjustment. AKRs have been reported in many plant species (26), including maize.

## Discussion

In plants, some linkage mapping studies based on small or medium population sizes have resulted in biased conclusions, detecting a small number of QTL (median, six) with surprisingly large PVE (≥50%) (27, 28). Different conclusions regarding the number, magnitude, and distribution of QTL for complex traits might be drawn if larger populations were evaluated (29, 30). Comparing LD mapping of 305 lines versus 522 lines demonstrates the beneficial effect of larger population size in this study (Table S6).

SNP assay is the test of choice for genomewide genetic analysis, but unfortunately it usually detects only two alleles at each locus (31, 32). This biallelic limitation can be overcome by using haplotypes constructed from multiple SNP markers likely to be within the same gene or from SNPs within 10 kb, the average size of LD blocks in maize. The shift to haplotype-based analysis substantially improved the efficiency of LD mapping. Compared with single SNP-based analysis, using haplotypes substantially improved the efficiency, with $P$ value decreasing from $5.8 \times 10^{-4}$ to $7.59 \times 10^{-8}$. Similarly, the accumulative PVE increased from 5.38% to 23.25% for the 522-line set. Combining single SNP– and haplotype-based analyses further increased the sum of PVE to 28.63% (Table S6). Given the expected >32,000 maize genes, a haplotype comprising 5–10 SNPs per gene is feasible within

a single experiment using high-density chips, allowing precise LD mapping and accelerating gene discovery.

Alleles occurring at low frequency in the mapping population are a major limiting factor for LD mapping. Even with SNP markers detecting only two alleles per locus, 277 (13.5%) of the SNP markers tested in this study had a minor allele frequency below 5%. Given that the number of individuals with a specific genotype can be very small, the effect of rare alleles on mapping could go far beyond the effect of small population sizes. Fortunately, biparental population–based mapping can use alleles that occur at low frequency in natural populations by designing crosses to create artificial populations with inflated frequencies of those alleles. As a result, 113 of the 277 SNP markers (40.8%) had a roughly equal ratio of both alleles in at least one of the RIL populations, allowing identification of three additional markers significantly associated with ASI.

We compared our mapping results with those from previous reports using the same RIL populations but tested in multiple seasons (33, 34). A total of 11 QTL were identified in nine genomic regions in a previous report using C5 (33). We confirmed the QTL for WW at bin 2.08 by linkage analysis. Around two of the regions (bins 5.06 and 8.03), we identified one QTL for WS (bin 5.07) and one QTL for WW (bin 8.02) by linkage mapping with the XB population. Using C6, a previous study identified a total of eight QTL in six genomic regions (34). One of the regions (bin 4.10) was confirmed by linkage analysis in the present study. QTL for WW were identified in the XB population in two of these regions (bins 1.02 and 8.02). Interestingly, for six of the nine regions that were not identified in the present study by linkage mapping with any of the RIL populations, a QTL at the same or a flanking bin has been identified by LD, integrated mapping, or both.

This study provides an example of joint linkage-LD mapping in maize. Parallel mapping, in which results were merged after independent linkage and LD analyses, provided complementary results by identifying common and specific QTL, whereas integrated mapping allowed simultaneous estimation of several genetic parameters. Linkage mapping of three RIL populations identified nine QTL, LD mapping with 305 lines identified only four QTL, and together the two methods (i.e., parallel mapping) identified 12 unique QTL. However, integrated mapping identified 18 additional QTL that were not identified by either linkage or LD mapping (Table S6). Integrated mapping also substantially improved the tests' statistical significance level, resulting in identification of two haplotype loci that contributed significantly to ASI. It could be even more efficient if multiple panels of mapping materials, including both biparental and natural populations, were collated. Using more inbred lines combined with different sets of biparental populations would be a particularly good scenario.

Plant breeding based on phenotypic selection under drought stress has resulted in substantial genetic gain (18), but this process has been slow, time-consuming, and erratic. The development of effective markers for drought resistance offers the promise of increasing the selective gain while also ensuring sustainable targeted breeding progress (35). A large-scale drought stress QTL mapping program conducted at the International Maize and Wheat Improvement Center (CIMMYT) over the last 12 years has identified >1,000 QTL associated with nine component traits, including ASI (20). However, no individual QTL was found to account for more than 20% of PVE, and most ranged from 4% to 10% of PVE. Thus, it has been widely concluded that no large-effect QTL existed for drought resistance or tolerance in maize. In this study, the identified genetic effect has been larger compared with results obtained in the linkage approach, and haplotype-based LD mapping allowed identification of single QTL with PVE up to 22.7%, whereas integrated mapping using haplotypes increased QTL PVE up to 34.7%. The sub-

stantial improvement seen after the shift to haplotypes was contributed by the multiple SNPs contained in the haplotype, which independently contributed only a small effect but collectively conferred a large effect.

As demonstrated by the present study, joint linkage–LD mapping (both parallel and integrated mapping) can overcome some constraints of genetic mapping methods and simultaneously open up opportunities to identify markers useful for breeding. This approach can be implemented in virtually all crops. Meanwhile, combining these analytical approaches with the use of haplotypes has allowed us to identify large-effect QTL for ASI in maize. The use of SNP haplotypes is likely to substantially improve QTL mapping power in many crops and also to facilitate large-scale allele mining in breeding programs.

## Materials and Methods

**Plant Materials.** A total of 522 maize lines were used in this study (Dataset S3). The materials used for linkage mapping and for joint linkage–LD mapping included selections from three RIL populations: XB (X178 × B73; n = 186), C5 (Ac7643 × Ac7729/TZSRW; n = 234), and C6 (CML444 × Malawi; n = 236). B73, Ac7643, and CML444 are drought-tolerant lines. Based on phenotypic data collected in a previous season, between 30 and 39 lines with extreme ASI phenotypes were chosen for selective genotyping, resulting in a total of 71 RILs from XB, 69 RILs from C5, and 77 RILs from C6 used for linkage mapping. The marker density, RIL population sizes, and tail population sizes (the individuals chosen from each tail) have been shown to have a 95% probability of detecting QTL with relatively small effects (5–10% of PVE) (36).

A total of 305 inbred lines were used for LD and joint linkage–LD mapping. This total included 105 introgression lines from the Chinese Academy of Agricultural Sciences that were developed via between two and four backcrosses with five recurrent parents and 30 donor parents, followed by one to three rounds of selfing (37).

**Field Experiment and Phenotyping.** All plant materials were tested under two water regimes (WW and WS) using an alpha (0, 1) lattice field design during the dry winter season (November 2007–April 2008) at the CIMMYT. The field design included two replications of each treatment. Each plot consisted of a 5-m row of 26 plants. All of the plants received the first three irrigations at 1, 18, and 41 d after planting. In the WW regime, irrigation was provided every 2 wk thereafter, but in the WS regime, irrigation was provided only at 65 d after planting, with no further irrigation until 50% anthesis occurred (at ~101 d after planting). ASI was scored as the difference (in days) between male and female flowering times. Selection of phenotypic extremes based on the ASI data collected in a previous season resulted in significant phenotypic difference but continuous variation in each RIL population.

**SNP Genotyping and Data Scoring.** Leaf samples were harvested from 15 plants of each line and bulked for extraction of total genomic DNA. All lines were genotyped using two 1,536-SNP chips via a GoldenGate assay (38). One chip was designed based on random candidate genes (RA chip), which were chosen with no prior knowledge or consideration of the function of the proteins (or RNAs) that they encode. This provides a random sample of genes across the genome for background control. The other chip (DT chip) was designed based on genes of known or suspected function that are likely involved in the control of drought response (39). The two chips provided 2,052 informative SNP markers of high quality (Dataset S1). Our sequence annotation analysis based on a Filtered Gene Set of maizesequence (http://www.maizesequence.org) indicated that there were 901 intragenic SNPs involving 781 genes among the 1,035 SNPs from the RA chip. Gene Ontology analysis and a literature review for the 1,017 SNPs from the DT chip identified 659 SNPs developed from 413 candidate genes that are likely involved in the pathways related to drought response (Dataset S2B). The DT chip allows candidate gene–based LD mapping for drought resistance. SNPs that were within 10-kb blocks were assigned as different alleles of a single haplotype locus.

**Data Analysis.** The maize germplasm used in this study was arranged into two line sets for analysis of population size effect and comparative mapping: the entire line set (n = 522) and an inbred line set (n = 305). PowerMarker version 3.25 (40) was used to calculate allele frequency and polymorphic information content. For each set of germplasm, principal component analysis was conducted to visualize the genetic structure, and pairwise relatedness coefficients (kinship matrix) were calculated using TASSEL 2.1 (5, 41).

QTL linkage mapping was implemented for each RIL population using WinQTL Cartographer version 2.5. For CIM, empirical thresholds for declaring significant QTL at $P < 0.05$ were determined by performing 1,000 permutations.

LD mapping was performed using mixed-linear model implemented with Tassel 2.1. The first five principal components derived from analysis of the $n = 522$ and $n = 305$ populations screened with the 1,035 RA chip markers were included in the mixed linear model analysis in place of the population structure (Q matrix). Both principal component and $K$ matrices were used, to minimize spurious associations.

The efficiency of integrated mapping was determined by the change in $P$ values when eight genetic models were applied to the entire line set ($n = 522$). In model A, the SNP allele has the same effect on ASI in three RIL populations and one inbred population. In model B, the allele effect in the three RIL populations is opposite to that in the inbred population. In models C, D, and E, the allele effect in one of the three RIL populations is opposite to that in the inbred population. In models F, G and H, the allele effect in two of the three RIL populations is opposite to that in the inbred population. Model A is equivalent to the LD mapping using the entire line set.

For both LD mapping and integrated linkage–LD mapping, the threshold for multiple tests to declare associations was determined by the Bonferroni correction (21). When performing $n$ tests, if the significance level for the entire series of tests is $\alpha$, then each of the tests should have probability $P = \alpha/n$. When the numbers of haplotype loci, markers, and markers by genetic models are 386, 2,052 and $2,052 \times 8$, respectively, at $\alpha = 0.01$, the Bonferroni-corrected thresholds for the $P$ value should be $2.59 \times 10^{-5}$, $4.87 \times 10^{-6}$, and $6.09 \times 10^{-7}$, with corresponding $-\log P$ values of 4.59, 5.31, and 6.22. A suggested level of significance was declared when SNP–trait associations were detected at $P < 0.001$ or $-\log P > 3$.

1. Myles S, et al. (2009) Association mapping: Critical considerations shift from genotyping to experimental design. *Plant Cell* 21:2194–2202.
2. Tuberosa R, Salvi S (2009) QTL for agronomic traits in maize production. *Maize Handbook, Volume I. Its Biology*, eds Bennetzen JL, Hakes S (Springer, New York), pp 501–541.
3. Buckler ES, 4th, (2002) Thornsberry JM (2002) Plant molecular diversity and applications to genomics. *Curr Opin Plant Biol* 5:107–111.
4. Yu J, Buckler ES (2006) Genetic association mapping and genome organization of maize. *Curr Opin Biotechnol* 17:155–160.
5. Yu J, et al. (2006) A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nat Genet* 38:203–208.
6. Devlin B, Roeder K (1999) Genomic control for association studies. *Biometrics* 55: 997–1004.
7. Zheng G, Freidlin B, Li Z, Gastwirth JL (2005) Genomic control for association studies under various genetic models. *Biometrics* 61:186–192.
8. Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* 155:945–959.
9. Falush D, Stephens M, Pritchard JK (2003) Inference of population structure using multilocus genotype data: Linked loci and correlated allele frequencies. *Genetics* 164: 1567–1587.
10. Zhu CS, Yu JM (2009) Nonmetric multidimensional scaling corrects for population structure in association mapping with different sample types. *Genetics* 182:875–888.
11. Harjes CE, et al. (2008) Natural genetic variation in lycopene epsilon cyclase tapped for maize biofortification. *Science* 319:330–333.
12. Buckler ES, et al. (2009) The genetic architecture of maize flowering time. *Science* 325: 714–718.
13. Maher B (2008) Personal genomes: The case of the missing heritability. *Nature* 456: 18–21.
14. Wu R, Zeng ZB (2001) Joint linkage and linkage disequilibrium mapping in natural populations. *Genetics* 157:899–909.
15. Meuwissen THE, Karlsen A, Lien S, Olsaker I, Goddard ME (2002) Fine mapping of a quantitative trait locus for twinning rate using combined linkage and linkage disequilibrium mapping. *Genetics* 161:373–379.
16. Wu R, Ma CX, Casella G (2002) Joint linkage and linkage disequilibrium mapping of quantitative trait loci in natural populations. *Genetics* 160:779–792.
17. Yu J, Holland JB, McMullen MD, Buckler ES (2008) Genetic design and statistical power of nested association mapping in maize. *Genetics* 178:539–551.
18. Bänziger M, Araus J (2007) Recent advances in breeding maize for drought and salinity stress tolerance. *Advances in Molecular Breeding Toward Drought and Salt-Tolerant Crops*, eds Jenks MA, Hasegawa PM, Jain SM (Springer, Dordrecht, The Netherlands), pp 587–601.
19. Edmeades GO, et al. (1993) Causes for silk delay in a lowland tropical maize population. *Crop Sci* 33:1029–1035.
20. Ribaut JM (2009) Drought tolerance in maize. *Handbook of Maize: Its Biology*, eds Bennetzen JL, Hake SC (Springer, New York), pp 311–344.
21. Abdi H (2007) Bonferroni and Šidák corrections for multiple comparisons. *Encyclopedia of Measurement and Statistics*, ed Salkind NJ (Sage, Thousand Oaks, CA).
22. Wilk MB, Gnanadesikan R (1968) Probability plotting methods for the analysis of data. *Biometrika* 55:1–17.
23. Krishnan A, et al. (2010) Genetic networks underlying plant abiotic stress responses. *Genes for Plant Abiotic Stress*, eds Matthew AJ, Wood AJ (Blackwell, Ames, IA), pp 263–279.
24. Ng DWK, et al. (2007) Plant SET domain-containing proteins: Structure, function and regulation. *Biochim Biophys Acta Gene Struc Expr* 1769:316–329.
25. Gábor V, et al. (2007) Engineering of reactive species detoxification pathways for increasing stress tolerance in plants. *Biotechnology and Sustainable Agriculture 2006 and Beyond*, eds Xu Z, et al. (Springer, Dordrecht, The Netherlands), pp 163–166.
26. Karuna Sree B, Rajendrakumar CSV, Reddy AR (2000) Aldose reductase in rice (*Oryza sativa* L.): Stress response and developmental specificity. *Plant Sci* 160:149–157.
27. Kearsey MJ, Farquhar AGL (1998) QTL analysis in plants: Where are we now? *Heredity* 80:137–142.
28. Xu Y (2002) Global view of QTL: Rice as a model. *Quantitative Genetics, Genomics and Plant Breeding*, ed Kang MS (CAB International, Wallingford, UK), pp 109–134.
29. Beavis WD (1998) QTL analyses: Power, precision and accuracy. *Molecular Dissection of Complex Traits*, ed Paterson AH (CRC, Boca Raton, FL), pp 145–162.
30. Schön CC, et al. (2004) Quantitative trait locus mapping based on resampling in a vast maize testcross experiment and its relevance to quantitative genetics for complex traits. *Genetics* 167:485–498.
31. Lu Y, et al. (2009) Molecular characterization of global maize breeding germplasm based on genome-wide single nucleotide polymorphisms. *Theor Appl Genet* 120: 93–115.
32. Yan J, et al. (2010) High-throughput SNP genotyping with the GoldenGate assay in maize. *Mol Breed* 25:441–451.
33. Welcker C, Boussuge B, Bencivenni C, Ribaut JM, Tardieu F (2007) Are source and sink strengths genetically linked in maize plants subjected to water deficit? A QTL study of the responses of leaf growth and of anthesis-silking interval to water deficit. *J Exp Bot* 58:339–349.
34. Messmer R, et al. (2009) Drought stress and tropical maize: QTL-by-environment interactions and stability of QTLs across environments for yield components and secondary traits. *Theor Appl Genet* 119:913–930.
35. Tuberosa R, Salvi S (2006) Genomics-based approaches to improve drought tolerance of crops. *Trends Plant Sci* 11:405–412.
36. Sun Y, et al. (2010) Efficiency of selective genotyping for genetic analysis of complex traits and potential applications in crop improvement. *Mol Breed* 26:493–511.
37. Hao Z, et al. (2009) Identification of quantitative trait loci for drought tolerance at seedling stage by screening a large number of introgression lines in maize. *Plant Breed* 128:337–341.
38. Fan JB, et al. (2006) Illumina universal bead arrays. *Methods Enzymol* 410:57–73.
39. Yan J, et al. (2009) Genetic characterization and linkage disequilibrium estimation of a global maize collection using SNP markers. *PLoS ONE* 4:e8451.
40. Liu K, Muse SV (2005) PowerMarker: An integrated analysis environment for genetic marker analysis. *Bioinformatics* 21:2128–2129.
41. Bradbury PJ, et al. (2007) TASSEL: Software for association mapping of complex traits in diverse samples. *Bioinformatics* 23:2633–2635.

# Supporting Information

## Lu et al. 10.1073/pnas.1006105107
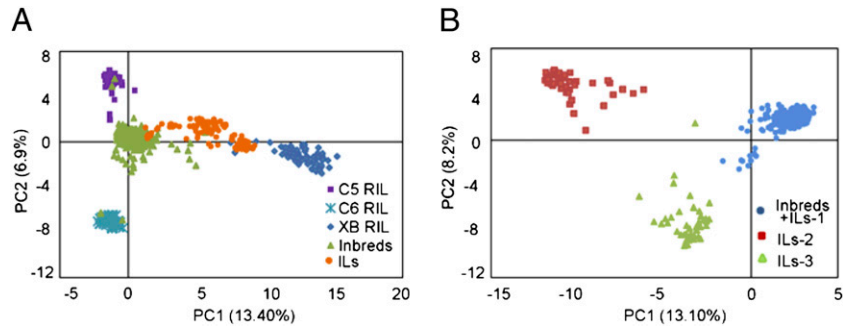


**Fig. S1.** Principal component analysis for the entire set of maize lines (*n* = 522) (*A*) and inbred lines + introgression lines (*n* = 305) (*B*).
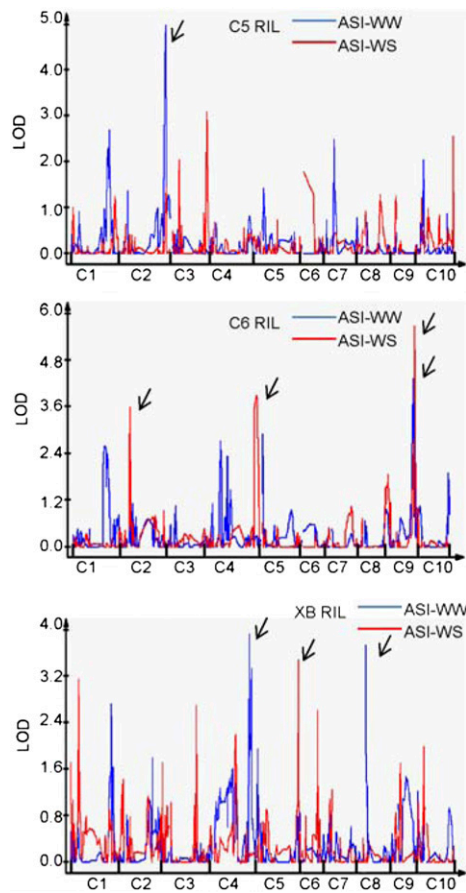


**Fig. S2.** Genome-wide scan of QTL for ASI under WW (blue lines) and WS (red lines) regimes using composite interval mapping and C5, C6, and XB RIL populations. Identified QTL are represented by arrows. C1–C10 represent chromosomes 1–10.
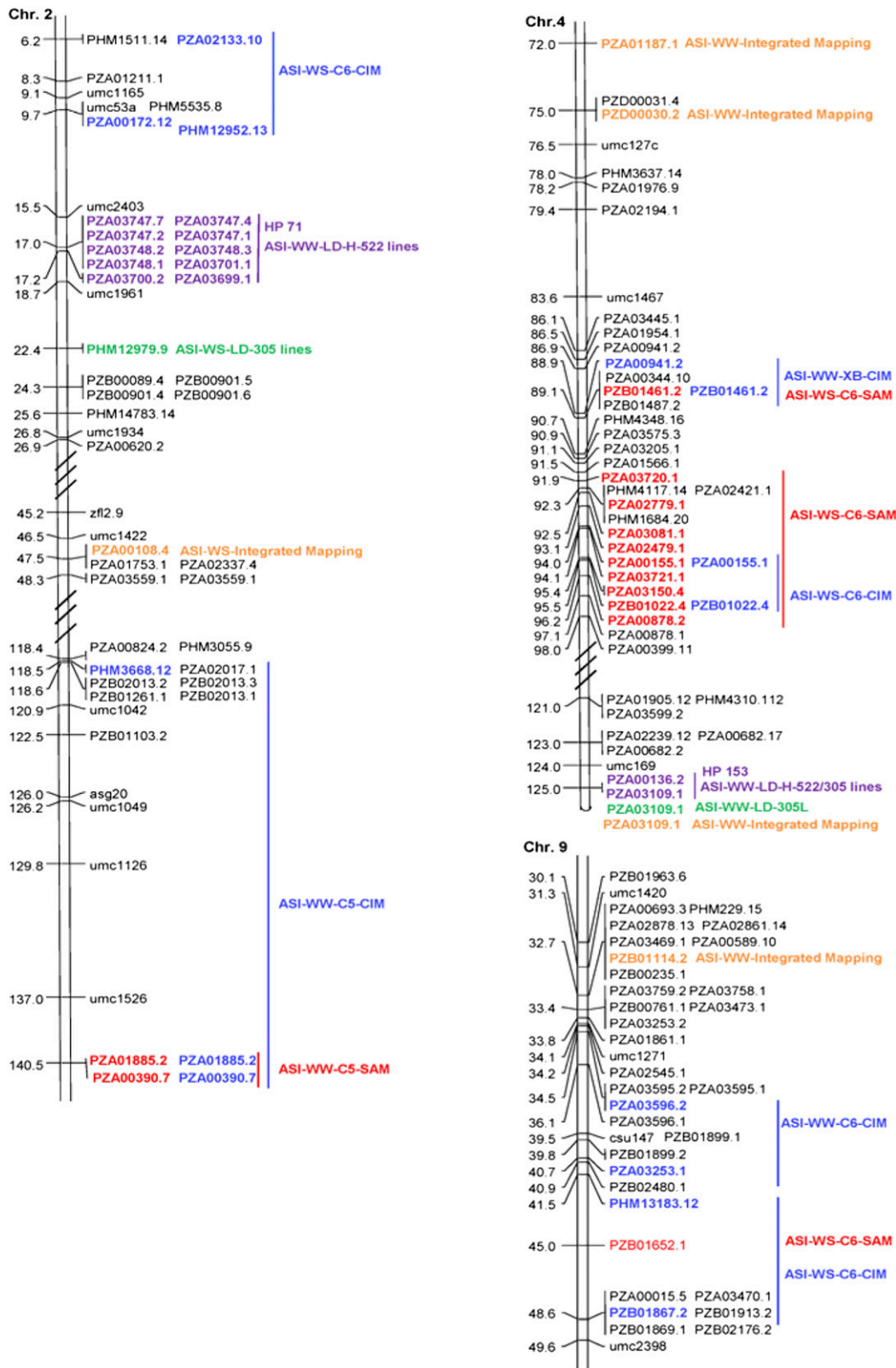
**Fig. S3.** QTL clusters identified for ASI under WW and WS regimes. QTL identified by different methods are distinguished by colors: linkage mapping via single-marker analysis in red and CIM in blue, using C5, C6, and XB RIL populations; LD mapping using single SNPs in green and using haplotypes in purple; and integrated linkage–LD mapping in orange.

**Table S1. Overview of alleles and SNPs in each haplotype locus from which two or more SNPs were amplified**

| No. of SNPs per haplotype locus | No. of haplotype loci | Number of haplotype loci with different numbers of alleles per haplotype locus | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 16 | 18 | 21 | 23 |
| 2 | 268 | 17 | 159 | 92 | | | | | | | | | | | | | | |
| 3 | 70 | | 10 | 26 | 20 | 10 | 3 | 1 | | | | | | | | | | |
| 4 | 25 | | 1 | 1 | 2 | 6 | 6 | 2 | 2 | 3 | 1 | 1 | | | | | | |
| 5 | 13 | | | 1 | 3 | 1 | 2 | 1 | 2 | | 1 | 2 | | | | | | |
| 6 | 4 | | | | 1 | | | 1 | | | 1 | | | | | | | 1 |
| 7 | 2 | | | | | | | | | 1 | | | | 1 | | | | |
| 8 | 1 | | | | | | | | | | | 1 | | | | | | |
| 9 | 1 | | | | | | | | ` | | | | | | 1 | | | |
| 10 | 1 | | | | | | | | | | | | | | | | 1 | |
| 12 | 1 | | | | | | | | | | | | | | | 1 | | |
| Total | 386 | 17 | 170 | 120 | 26 | 17 | 11 | 5 | 4 | 4 | 3 | 3 | 1 | 1 | 1 | 1 | 1 | 1 |

**Table S2. Summary information for haplotypes across the maize genome**

| Chr | No. of haplotype loci | No. of SNPs involved | No. of SNPs per locus | Average Interval distance between two haplotype loci, Mb |
|---|---|---|---|---|
| 1 | 64 | 164 | 2.5625 | 4.7725 |
| 2 | 42 | 119 | 2.8333 | 5.6741 |
| 3 | 46 | 114 | 2.4783 | 5.0310 |
| 4 | 43 | 102 | 2.3721 | 5.7693 |
| 5 | 45 | 112 | 2.4889 | 4.8132 |
| 6 | 22 | 53 | 2.4091 | 7.8454 |
| 7 | 38 | 88 | 2.3158 | 4.5106 |
| 8 | 37 | 101 | 2.7297 | 4.4427 |
| 9 | 28 | 72 | 2.5714 | 5.1935 |
| 10 | 21 | 63 | 3.0000 | 7.3423 |
| Overall | 386 | 988 | 2.5596 | 5.5395 |

**Table S3. Marker–ASI associations identified by single marker analysis using three RIL populations**

| QTL | SNP | Trait | Population | Chr | Bin | LOD | Additive effect |
|---|---|---|---|---|---|---|---|
| 1 | umc13.1 | ASI-WS | XB | 1 | 1.04 | 2.83 | 3.508 |
| 1 | PZB00872.3 | ASI-WS | XB | 1 | 1.04 | 2.82 | 3.51 |
| 1 | PZA00081.18 | ASI-WS | XB | 1 | 1.04 | 2.84 | 3.501 |
| 1 | PZA03183.5 | ASI-WS | XB | 1 | 1.04 | 2.83 | 3.508 |
| 2 | PZA01885.2 | ASI-WW | C5 | 2 | 2.09 | 3.10 | −1.873 |
| 2 | PZA00390.7 | ASI-WW | C5 | 2 | 2.09 | 3.09 | −1.883 |
| 3 | PZB01461.2 | ASI-WS | C6 | 4 | 4.09 | 2.51 | −1.983 |
| 4 | PZA03721.1 | ASI-WS | C6 | 4 | 4.09 | 2.64 | −1.983 |
| 4 | PZA03720.1 | ASI-WS | C6 | 4 | 4.09 | 2.64 | −1.983 |
| 4 | PZA03081.1 | ASI-WS | C6 | 4 | 4.10 | 2.65 | −2.01 |
| 4 | PZA02479.1 | ASI-WS | C6 | 4 | 4.10 | 3.00 | −2.127 |
| 4 | PZA00878.2 | ASI-WS | C6 | 4 | 4.10 | 3.21 | −2.179 |
| 4 | PZA02779.1 | ASI-WS | C6 | 4 | 4.09 | 2.52 | −1.983 |
| 4 | PZA00155.1 | ASI-WS | C6 | 4 | 4.10 | 3.17 | −2.146 |
| 4 | PZB01022.4 | ASI-WS | C6 | 4 | 4.10 | 3.24 | −2.173 |
| 4 | PZA03150.4 | ASI-WS | C6 | 4 | 4.10 | 3.18 | −2.145 |
| 5 | PZB01652.1 | ASI-WS | C6 | 9 | 9.05 | 2.92 | −2.021 |

**Table S4. Marker–ASI associations identified by composite interval mapping using three RIL populations**

| Trait | Left SNP | Peak SNP | Right SNP | Population | LOD threshold | LOD score | Chr | Bin | $R^2$, % | Additive effect |
|-------|----------|----------|-----------|------------|---------------|-----------|-----|-----|----------|-----------------|
| ASI-WW | PHM3668.12 | PZA01885.2 | PZA00390.7 | C5 | 3.5 | 5.0 | 2 | 2.09 | 20.1 | −1.94 |
| | PZA03596.2 | PZB01899.2 | PZA03253.1 | C6 | 4.3 | 4.3 | 9 | 9.05 | 18.2 | −1.7 |
| | PZA00941.2 | PZA00344.10 | PZB01461.2 | XB | 3.5 | 3.9 | 4 | 4.09 | 21.9 | −3 |
| | PZA02174.2 | PZA01623.3 | PZA00058.1 | | | 3.7 | 8 | 8.02 | 20.7 | 2.38 |
| ASI-WS | PZA02133.10 | PZA00172.12 | PHM12952.13 | C6 | 3.0 | 3.6 | 2 | 2.03 | 14.8 | 2.23 |
| | PZA00155.1 | PZA03150.4 | PZB01022.4 | | | 3.9 | 4 | 4.10 | 16.2 | −1.81 |
| | PHM13183.12 | PZB01652.1 | PZB01867.2 | | | 5.7 | 9 | 9.05 | 25.8 | −2.3 |
| | PZA02068.1 | PZA00545.26 | PZA02015.11 | XB | 3.2 | 3.5 | 5 | 5.07 | 17.5 | −2.53 |

**Table S5. Comparison of QTL identified for ASI under WW and WS regimes by single SNP- and haplotype-based LD mapping**

| | | Type of genotype data | | | | | |
|---|---|---|---|---|---|---|---|
| | | Single SNP | | Haplotype | | Single SNP + haplotype | |
| Trait | Parameter | $n = 305$ | $n = 522$ | $n = 305$ | $n = 522$ | $n = 305$ | $n = 522$ |
| ASI-WW | Number of QTL | 3 | 4 | 2 | 2 | 5 | 6 |
| | −Log ($P$) | 3.88 | 3.29 | 3.89 | 5.95 | 3.89 | 4.62 |
| | Average of $R^2$, % | 2.34 | 1.54 | 3.94 | 12.50 | 3.14 | 7.02 |
| | Sum of $R^2$, % | 7.02 | 6.16 | 7.87 | 25.00 | 14.89 | 31.16 |
| ASI-WS | Number of QTL | 1 | 3 | 0 | 1 | 1 | 4 |
| | −Log ($P$) | 3.14 | 3.18 | — | 8.30 | 3.14 | 5.74 |
| | Average of $R^2$, % | 1.89 | 1.53 | — | 21.50 | 1.89 | 11.52 |
| | Sum of $R^2$, % | 1.89 | 4.60 | — | 21.50 | 1.89 | 26.10 |
| Mean | Number of QTL | 2.00 | 3.50 | 1.00 | 1.50 | 3.00 | 5.00 |
| | −Log ($P$) | 3.51 | 3.24 | 3.89 | 7.12 | 3.89 | 5.18 |
| | Average of $R^2$, % | 2.12 | 1.54 | 3.94 | 17.00 | 3.14 | 9.27 |
| | Sum of $R^2$, % | 4.46 | 5.38 | 7.87 | 23.25 | 8.39 | 28.63 |

For single SNP-based LD mapping, "Bonferroni-corrected" threshold for multiple tests is −log $P$ > 5.31. For haplotype-based linkage LD mapping, the threshold is −log $P$ > 4.59. SNP–trait associations were included for comparison at the suggestive significance level with −log $P$ > 3.00. Results for LD mapping with 522 lines are from model A. — indicates no data available.

**Table S6. Comparison of single SNP-based integrated linkage–LD mapping with linkage mapping and LD mapping**

| Trait | Marker | Chr | Bin | Position | SNP* | LD mapping −Log P | R², % | Allele effect | Integrated mapping −Log P | R², % | Allele effect | Linkage mapping RIL | LOD | R², % | Allele effect |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| WW | PHM1438.34 | 1 | 1.08 | 212389447 | **A**/G | 4.66 | 2.74 | −0.357 | 3.08 | 1.41 | −0.226 | C5 | 1.657 | 6.64 | −1.143 |
| WW | PZA03109.1 | 4 | 4.02 | 2639015 | C/**G** | 3.65 | 2.12 | −0.276 | 3.65 | 2.12 | −0.314 | C5 | 0.363 | 1.64 | −1.036 |
| WW | PZA02792.25 | 5 | 5.04 | 21771014 | **A**/C | 3.67 | 2.16 | 0.348 | 3.04 | 1.38 | 0.252 | No genetic or background effect | | | |
| WW | PZA02135.2 | 1 | 1.06 | 166550294 | **A**/G | 1.19 | 0.54 | 0.152 | 3.78 | 1.78 | 0.247 | C6 | 1.913 | 8.19 | −1.158 |
| | | | | | | | | | | | | C5 | 0.587 | 2.02 | 0.730 |
| WW | PZA01713.4 | 4 | 4.06 | 35384247 | **C**/G | 1.92 | 0.99 | 0.217 | 3.59 | 1.71 | 0.226 | XB | 1.238 | 5.24 | 1.110 |
| WW | PZB01689.2 | 4 | 4.06 | 129831472 | A/**G** | 2.57 | 1.66 | −0.229 | 3.37 | 1.73 | −0.246 | C5 | 0.187 | 0.62 | 0.378 |
| WW | PZA01187.1 | 4 | 4.08 | 177666738 | **A**/C | 0.93 | 0.37 | 0.132 | 3.18 | 1.41 | 0.222 | C6 | 0.268 | 1.02 | −0.510 |
| WW | PZD00030.2 | 4 | 4.08 | 177976082 | **A**/G | 0.77 | 0.28 | 0.114 | 3.21 | 1.41 | 0.235 | C6 | 0.323 | 1.30 | −0.585 |
| | | | | | | | | | | | | XB | 0.115 | 0.54 | 0.439 |
| WW | PZA01144.1 | 6 | 6.06 | 147820659 | **A**/G | 1.64 | 0.8 | −0.314 | 3.33 | 1.53 | −0.395 | XB | 0.672 | 2.99 | −0.819 |
| WW | PZB01388.1 | 6 | 6.08 | 161663864 | **C**/G | 2.32 | 1.32 | −0.195 | 3.10 | 1.46 | −0.211 | XB | 0.822 | 3.55 | −0.899 |
| WW | PZB01114.2 | 9 | 9.04 | 60776186 | **C**/G | 2.31 | 1.17 | −0.264 | 3.08 | 1.37 | −0.234 | C6 | 0.216 | 0.75 | 0.457 |
| WS | PZB02144.5 | 1 | 1.09 | 238902255 | **A**/T | 2.01 | 1.16 | 0.339 | 3.39 | 1.62 | 0.413 | XB | 0.206 | 0.79 | 0.611 |
| WS | PZA00108.4 | 2 | 2.03 | 13779970 | **A**/T | 2.24 | 1.36 | −0.268 | 3.36 | 1.66 | −0.230 | C6 | 0.688 | 3.33 | −0.854 |
| | | | | | | | | | | | | XB | 0.110 | 0.36 | 0.406 |
| WS | PZA03535.3 | 4 | 4.06 | 107751627 | A/**C** | 1.94 | 1.23 | 0.572 | 3.13 | 1.55 | 0.576 | XB | 0.117 | 0.38 | 0.413 |
| WS | PZA02148.1 | 6 | 6.06 | 143875945 | **A**/G | 1.55 | 0.84 | 0.230 | 3.01 | 1.43 | 0.271 | C5 | 0.145 | 0.74 | 0.371 |
| | | | | | | | | | | | | XB | 0.155 | 0.50 | 0.464 |
| WW | PZA00726.8 | 4 | 4.06 | 60767943 | **A**/G | 2.29 | 1.19 | 0.285 | 3.16 | 1.41 | 0.266 | Opposite background effect from C6 | | | |
| WW | PZA01039.1 | — | — | — | **A**/G | 2.36 | 1.27 | −0.308 | 3.49 | 1.65 | −0.281 | Opposite background effect from C6 and XB | | | |
| WW | PZA00398.4 | 7 | 7.03 | 15154697 | **A**/G | 1.63 | 0.76 | 0.152 | 3.32 | 1.47 | 0.190 | Opposite background effect from all RILs | | | |
| WS | PZA03579.2 | 8 | 8.04 | 105795595 | **A**/G | 2.49 | 1.55 | 0.247 | 3.21 | 1.52 | 0.255 | Opposite background effect from C5 and XB | | | |
| WS | PHM4134.8 | 8 | 8.04 | 105795742 | **C**/G | 2.60 | 1.74 | −0.287 | 3.28 | 1.63 | −0.280 | Opposite background effect from C5 and XB | | | |
| WS | PZA00286.1 | 3 | 3.05 | 65074690 | **A**/C | 0.61 | 0.23 | 0.125 | 3.82 | 1.78 | 0.311 | Opposite background effect from C5 and XB | | | |
| WS | PHM12979.9 | 2 | 2.03 | 9037148 | **A**/G | 3.14 | 1.89 | 0.260 | 1.02 | 0.34 | 0.094 | Unknown background effect | | | |
| WS | PZA00081.18 | 1 | 1.04 | 45675182 | **C**/G | 0.40 | 0.11 | −0.396 | 2.03 | 0.88 | 0.023 | XB | 2.84 | 10.99 | 3.501 |
| WW | PZA00390.7 | 2 | 2.09 | 206974778 | A/**G** | 0.24 | 0.05 | −0.058 | 1.93 | 0.79 | 0.174 | C5 | 3.09 | 13.10 | −1.883 |
| WS | PZA03081.1 | 4 | 4.10 | 215393350 | **C**/G | 2.62 | 1.6 | 0.516 | 2.30 | 1.03 | 0.036 | C6 | 2.65 | 9.00 | −2.010 |
| WS | PZA02479.1 | 4 | 4.10 | 218367682 | **A**/G | 1.37 | 0.72 | −0.235 | 1.78 | 0.75 | 0.017 | C6 | 3.00 | 12.89 | 2.127 |
| WS | PZB01652.1 | 9 | 9.05 | 106850855 | **A**/T | 0.21 | 0.04 | −0.054 | 1.44 | 0.56 | −0.072 | C6 | 2.92 | 9.04 | 2.021 |

For LD mapping, $n = 2,052$, and the Bonferroni-corrected threshold for multiple tests is −log $P > 5.31$. For integrated linkage–LD mapping, $n = 2,052 \times 8$, and the threshold is −log $P > 6.22$. The level for suggestive significance at −log $P > 3.00$ was used for comparison.
*Nucleotides in bold type have the allele effect in LD mapping.— indicates no data available.

## Other Supporting Information Files

Dataset S1 (XLSX)
Dataset S2 (XLSX)
Dataset S3 (XLSX)