nature genetics

Genome-wide association study of leaf architecture in the maize nested association mapping population

Feng Tian[1,11], Peter J Bradbury[2,11], Patrick J Brown[1,3], Hsiaoyi Hung[4], Qi Sun[5], Sherry Flint-Garcia[6,7], Torbert R Rocheford[3,8], Michael D McMullen[6,7], James B Holland[4,9] & Edward S Buckler[1,2,10]

US maize yield has increased eight-fold in the past 80 years, with half of the gain attributed to selection by breeders. During this time, changes in maize leaf angle and size have altered plant architecture, allowing more efficient light capture as planting density has increased. Through a genome-wide association study (GWAS) of the maize nested association mapping panel, we determined the genetic basis of important leaf architecture traits and identified some of the key genes. Overall, we demonstrate that the genetic architecture of the leaf traits is dominated by small effects, with little epistasis, environmental interaction or pleiotropy. In particular, GWAS results show that variations at the *liguleless* genes have contributed to more upright leaves. These results demonstrate that the use of GWAS with specially designed mapping populations is effective in uncovering the basis of key agronomic traits.

In the past century, continuing gains in grain yield of US corn have not been due to more grain per plant, but rather to adaptation of hybrids to continually higher plant densities[1]. The average plant density of maize in the United States has increased from 30,000 plants per hectare in the 1930s to >80,000 plants per hectare currently[1]. During the same period, the leaves of maize hybrids have become more upright[1]. Increasing the angle between the leaf midrib and the ground maintains light capture under high density[2,3]. Hybrids with vertically oriented leaves have considerable yield advantages in both model simulation[4] and field experiments[5]. A study in rice has also shown that more upright leaves not only improve light capture but also improve the accumulation of leaf nitrogen for grain filling[6]. Leaf angle, together with leaf size, are important components of leaf architecture, influencing canopy morphology and photosynthetic efficiency and, as a result, overall yield. Clarifying the genetic architecture of these traits will impact trait manipulation for continued maize improvement.

GWAS in diverse maize is challenging, as linkage disequilibrium decays within 2,000 bp[7]. However, with the recent discovery of 1.6 million SNPs by the maize HapMap project[7] and the development of a large joint linkage–association panel known as the nested association mapping (NAM) population[8], effective GWAS in maize have now become possible. The maize NAM panel was created by crossing 25 diverse lines of maize to one reference line, then producing ~5,000 recombinant inbred lines (RILs)[8].

Maize NAM offers several unique advantages for GWAS. (i) Population structure is controlled by the NAM design itself. By reshuffling the genomes of founders during RIL development, population structure within each population is eliminated. Although the difference between populations resulting from the diverse founders remains, this difference is controlled by including a population term in the GWAS model. (ii) Because maize HapMap SNPs are accurately imputed in the offspring based on parental genotypes, the reduced linkage disequilibrium in the parents owing to historical recombination and mutation can be used to improve mapping resolution. (iii) Phenotypic traits are accurately measured by growing ~50,000 plots with ~15 plants each over 9 environments. (iv) The joint linkage design allows ~90% of the background genetic variance to be controlled during GWAS. (v) Joint linkage and GWAS results can be directly compared.

We used this new maize NAM design to dissect the genetic architecture of upper leaf angle, leaf length and width. Using joint stepwise regression[9] (**Supplementary Note**), we identified 30–36 quantitative trait loci (QTLs) for the three leaf traits, explaining 74.8%–80.3% of the phenotypic variation and >83% of genetic variance (**Table 1**). Small-effect alleles underlie all QTLs for leaf traits (**Supplementary Fig. 1**). Similar genetic architectures have been observed for other complex maize traits[9,10], as well as in animals[11], in contrast with the

Table 1 QTLs identified by joint linkage mapping

Trait	Number of QTLs	Phenotypic variance explained (%)	Broad-sense heritability (%)
Leaf length (mm)	36	77.7	92.7
Leaf width (mm)	34	80.3	92.4
Upper leaf angle (°)	30	74.8	89.4

[1]Institute for Genomic Diversity, Cornell University, Ithaca, New York, USA. [2]US Department of Agriculture–Agricultural Research Service (USDA-ARS), Ithaca, New York, USA. [3]Department of Crop Sciences, University of Illinois, Urbana, Illinois, USA. [4]Department of Crop Science, North Carolina State University, Raleigh, North Carolina, USA. [5]Computational Biology Service Unit, Cornell University, Ithaca, New York, USA. [6]USDA-ARS, Columbia, Missouri, USA. [7]Division of Plant Sciences, University of Missouri, Columbia, Missouri, USA. [8]Department of Agronomy, Purdue University, Urbana, Illinois, USA. [9]USDA-ARS, Raleigh, North Carolina, USA. [10]Department of Plant Breeding and Genetics, Cornell University, Ithaca, New York, USA. [11]These authors contributed equally to this work. Correspondence should be addressed to E.S.B. (esb33@cornell.edu).

Received 2 June 2010; accepted 15 December 2010; published online 9 January 2011; doi:10.1038/ng.746

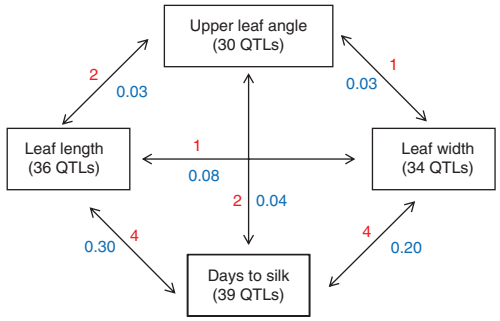

Figure 1 Low genetic overlap between leaf traits. The correlations of allelic effect estimates across founders for each pair of traits at colocalized QTLs were used to test the genetic overlap between different traits. Number in red, number of shared QTLs between traits. Number in blue, phenotypic correlation in r^2 between traits.

larger effects observed in self-pollinated plant species[12–14]. These results suggest that the genetic architecture of complex traits in outcrossing and selfing species may have evolved differently.

Although 11%–50% of QTLs for leaf traits showed considerable QTL by environment interactions after sequential Bonferroni correction[15], their effects are much smaller than QTL main effects (**Supplementary Note** and **Supplementary Tables 1–3**). No considerable epistasis was detected across and within populations for these traits after sequential Bonferroni correction[15] (see **Supplementary Note**, which discusses the power issues involved in detecting epistasis). Furthermore, a joint linkage additive QTL model can accurately predict the phenotypes of NAM parents (**Supplementary Fig. 2**), further indicating that epistatic effects were relatively unimportant in determining phenotype.

The NAM design provides an opportunity to test pleiotropy, the genetic overlap between different complex traits (Online Methods). We observed low pleiotropy among different leaf traits, explaining the weak phenotypic correlations among them (**Fig. 1**). Therefore, different sets of genetic variants probably control the natural variation in these leaf traits; this facilitates the pyramiding of favorable alleles for different traits in selective breeding.

Joint linkage mapping with 1,106 markers produced QTL support intervals ($P < 0.05$) that averaged several centimorgans (cM) (**Supplementary Tables 1–3**). To further dissect these QTLs, we conducted a GWAS by imputing 1.6 million HapMap SNPs discovered on the 27 NAM parents[7] onto offspring RILs. As a result, the historical recombination captured by HapMap SNPs was used to improve mapping resolution. HapMap SNPs with missing data were first imputed in the parents, and then all SNPs were imputed onto the RILs based on a procedure described in Online Methods.

We first tested NAM-GWAS on two simple traits with a known genetic basis. Across a 3.4 cM (5.2 Mb) region flanking the *p1* gene for cob color[16], which is covered by 4,879 SNPs, 4 SNPs spanning 492 kb showed the same genotypic pattern as cob color segregation in founders. Owing to the presence of >300 informative recombinants in the NAM population, NAM-GWAS identified the most significant of these SNPs, located in the tandem repeats of *p1* gene[16]. Across a 1 cM region (4.6 Mb) flanking the *y1* gene for kernel color[17], 1 SNP of 2,110 showed the genotypic pattern that matched kernel color segregation in the founders. NAM-GWAS identified this SNP, which was located 113 bp upstream of an insertion at *y1* (ref. 17) that is probably causative. These results showed that, as long as causal SNPs or SNPs in high linkage disequilibrium exist in the data set for large effect loci, NAM-GWAS can easily identify them from thousands of SNPs bracketed by flanking linkage markers.

For leaf quantitative traits, we used a subsampling-based multiple SNP model to carry out association mapping (Online Methods) based on a method successfully applied to GWAS in mice[18,19] and confirmed to be more robust than single-locus scanning[20]. Briefly, a subsample was formed by randomly sampling 80% of the lines from each population without replacement, and then forward regression was used to fit SNPs to a phenotype. The phenotypes used here were the residuals for each chromosome calculated from a joint linkage model. We conducted association analysis in 100 subsamples to sample the model space and measure the reliability of association position estimates. We used the bootstrap posterior probability (BPP), the proportion of the 100 models in which a SNP was included, to measure the robustness of SNP association. By using a permutation test, we found a threshold BPP of 0.05 led to a type I error rate <0.05 (Online Methods).

Out of 1.6 million tested SNPs, we detected 203, 287 and 295 significant SNPs with BPP ≥ 0.05 for upper leaf angle, leaf length and width, respectively (**Figs. 2** and **3**, **Supplementary Figs. 3** and **4** and **Supplementary Tables 4–6**). GWAS results overlapped significantly ($P < 0.05$) with joint linkage QTL but were not identical (**Supplementary Fig. 5**). For several reasons, we did not expect the results of joint linkage mapping and GWAS to be exactly the same. Instead of testing marker effects within a family, GWAS tests marker effects across families, creating larger allele classes and consequently making it easier to detect QTLs with smaller effects. With linkage mapping, neighboring QTLs or multiple alleles at a QTL can lead to a fused linkage signal owing to limited recent recombination during RIL development, whereas GWAS can potentially dissect the fused linkage signal into separate components by using the historical recombination present in parents. The fact that the set of 1.6 million SNPs is insufficient to capture all the haplotypes present in maize[7] probably leads to missed QTLs. Additionally, multiple rare causal variants that may be separated by relatively large genomic distances can create synthetic associations[21].

We further evaluated the proximity of significant SNPs associated with maize leaf length and width to candidate genes or microRNAs known to affect leaf shape in *Arabidopsis thaliana* and of SNPs significantly associated with leaf angle to genes known to affect the leaf blade-sheath boundary in maize (**Supplementary Note**). Although we observed no significant enrichment of association around candidate genes for leaf length and width (**Supplementary Table 7**), three microRNA genes (**Supplementary Table 8**) and a target gene of these microRNAs (**Supplementary Table 9**) showed significant enrichments. For upper leaf angle, we detected marked associations around the *lg1* (*liguleless1*) and *lg2* (*liguleless2*) genes (**Supplementary Table 10**), genes that have mutants known to affect leaf angle in maize[22,23].

The ligule and auricle are the regions separating the blade and sheath of a maize leaf, which allow the leaf blade to bend away from the stem. *lg1* and *lg2* mutants have no ligule or auricle, leading

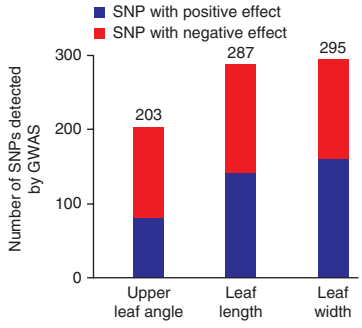

Figure 2 Summary of NAM-GWAS results for three leaf architecture traits.

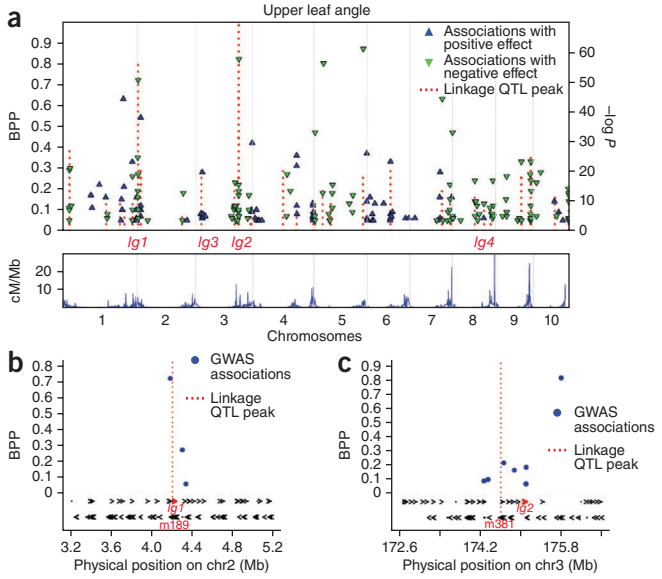

Figure 3 Associations around *lg1* and *lg2* explained the two most significant QTLs for upper leaf angle. (**a**) Overview of GWAS results for upper leaf angle. Red dotted lines, joint linkage QTL peaks; scale, $-\log_{10}$ of *P* value of QTLs. Triangles, associations identified by GWAS; scale, BPP. Four *liguleless* genes are indicated based on their physical position. Bottom, recombination rate along the chromosome. (**b**) Associations identified by GWAS at the *lg1* locus. chr2, chromosome 2. (**c**) Associations identified by GWAS at the *lg2* locus. Arrows pointing right and left, annotated genes on the positive- and negative-sense strands of DNA. chr3, chromosome 3.

to considerably more upright leaves than their normal counterparts[22,23]. *lg2* mutant alleles lead to significant grain yield increase in maize hybrids[2,3] (**Supplementary Fig. 6**), whereas the effect of the *lg1* mutant alleles depends on genetic background[3] (**Supplementary Fig. 6**).

The associations around *lg1* and *lg2* explained the two most significant QTLs for upper leaf angle (**Fig. 3a**). At the QTL near marker m189 (4.2 Mb) on chromosome 2, two SNPs jointly captured most of the association signals and spanned a physical region of 120 kb containing *lg1* (**Fig. 3b** and **Supplementary Table 11**). Furthermore, of 100 subsample models, 93% fitted exactly one of the two most significant SNPs but not both (**Supplementary Table 12**), suggesting that a single allele of major effect underlies the QTL at *lg1*. At the QTL near marker m381 (174.6 Mb) on chromosome 3, we detected a large cluster of associations spanning a physical region of 1.5 Mb around *lg2* (**Fig. 3c** and **Supplementary Table 11**). Seventy-three percent of the subsample models included two SNPs from this cluster, suggesting that two alleles seem to be responsible for the phenotypic variation in this region (**Supplementary Table 12**). Three of those SNPs are located in *lg2* introns (**Supplementary Table 11**), although the most significant SNP, with a BPP of 0.82, was ~685 kb downstream. The strongest association at the *lg1* locus is also located outside of *lg1*. One possible explanation in both cases is that the set of 1.6 million SNPs does not contain a causal polymorphism for the alleles involved and that the SNPs identified are in linkage disequilibrium with the causal polymorphisms. Alternatively, the variation controlling expression may be located tens of kilobases[24,25] to as far as 1 Mb[26] from the gene. In addition to the SNP associations near the *lg1* and *lg2* loci, weaker associations exist with the *lg4* locus, but none exist with the *lg3* locus (**Fig. 3a** and **Supplementary Table 11**).

To further examine the associations with the *liguleless* genes, we genotyped seven SNPs identified from GWAS in a panel of 282 diverse lines[27]. Four of the SNPs were weakly associated with upper leaf angle ($P < 0.1$) (**Supplementary Table 11**). This result is consistent with our NAM-GWAS results and with the relatively low power of an association panel of 282 lines to detect effects of <0.2 phenotypic standard deviations (s.d.) as shown by simulation (**Supplementary Fig. 7**).

The relative importance of common versus rare variants is critical to understanding the basis of natural variation and its application in medicine, agriculture and other areas of biology[28]. Using results from both joint linkage analysis and GWAS to address this question, we can compare the difference in allele frequency distribution between QTL and GWAS quantitative trait nucleotide (QTN) (**Fig. 4**). The number of NAM families containing non-B73 alleles of individual QTL tends to fall in the middle of the range, whereas the number of families containing individual GWAS QTN tends toward the lower end of the range. This observation reinforces the hypothesis that common QTL regions result from multiple underlying polymorphisms (**Fig. 4**). However, the allele frequency of GWAS QTN still differs from that of random SNPs, suggesting that the GWAS QTNs tend to be shared among lines. However, because QTNs and QTLs with very small effects remain undetected and because the set of 1.6 million SNPs probably does not contain most of the causative polymorphisms, the distributions observed here are probably different from the underlying true distributions.

We note some limitations of this NAM-GWAS. First, the set of 1.6 million SNPs is unlikely to capture all of the haplotypes present in the diverse maize inbred lines[7]. Second, only SNPs and small insertion-deletions are considered. However, the extensive structural variation in the form of copy number variation (CNV) and presence-absence variation (PAV) among maize inbreds may have functional importance and thus contribute to the variation that is not captured by SNPs[29,30]. Third, incorrect imputation of HapMap SNPs in NAM parents could lead to less accurate estimates of associations. Fourth, the moderate number of founders has limited resolution to dissect low-frequency QTN.

Despite these current limitations, we could use NAM-GWAS to identify individual genes controlling complex agronomic traits. The associations identified here provide a basis for further GWAS and fine-mapping efforts to pinpoint causal variants and to clarify how the implicated genes affect leaf architecture traits. These insights into the genetic basis of maize leaf architecture can now be used to further improve global maize with an optimum leaf architecture designed for high planting densities.

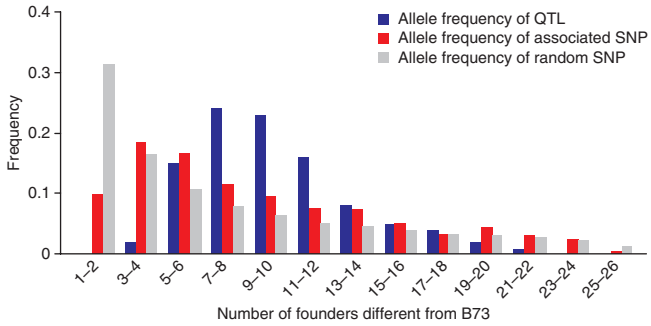

Figure 4 Comparison of allele frequency distributions of QTLs, associated SNPs and random SNPs. For comparison, a random SNP allele frequency distribution was produced by sampling 1.6 million HapMap SNPs. Both the allele frequency distributions of QTL and GWAS QTN differ from the random SNP allele frequency distribution. Notably, the associated SNPs show lower allele frequencies than QTL.

URLs. KBiosciences, http://kbioscience.co.uk/.

METHODS

Methods and any associated references are available in the online version of the paper at http://www.nature.com/naturegenetics/.

Note: Supplementary information is available on the Nature Genetics website.

ACKNOWLEDGMENTS

We thank L. Rigamer Lirette and S. Myles for editing the manuscript. This work was supported by US National Science Foundation grants (DBI-0820619, 0321467, 0703908 and 0638566) and USDA-ARS.

AUTHOR CONTRIBUTIONS

F.T. and P.J. Bradbury contributed equally to this work. M.D.M., J.B.H. and E.S.B. contributed to the study design. S.F.-G., T.R.R., M.D.M., J.B.H. and E.S.B. collected phenotypes. F.T., P.J. Bradbury, P.J. Brown, H.H., Q.S. and E.S.B. carried out analysis. F.T., P.J. Bradbury and E.S.B. wrote the paper. All authors discussed the results and commented on the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Published online at http://www.nature.com/naturegenetics/.
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/.

1. Duvick, D.N. Genetic progress in yield of United States maize (*Zea mays* L.). *Maydica* **50**, 193–202 (2005).
2. Pendleton, J.W., Smith, G.E., Winter, S.R. & Johnston, T.J. Field investigation of the relationships of leaf angle in corn (*Zea mays* L.) to grain yield and apparent photosynthesis. *Agron. J.* **60**, 422–424 (1968).
3. Lambert, R.J. & Johnson, R.R. Leaf angle, tassel morphology, and the performance of maize hybrids. *Crop Sci.* **18**, 499–502 (1978).
4. Duncan, W.G. Leaf angle, leaf area, and canopy photosynthesis. *Crop Sci.* **11**, 482–485 (1971).
5. Pepper, G.E., Pearce, R.B. & Mock, J.J. Leaf orientation and yield of maize. *Crop Sci.* **17**, 883–886 (1977).
6. Sinclair, T.R. & Sheehy, J.E. Erect leaves and photosynthesis in rice. *Science* **283**, 1455 (1999).
7. Gore, M.A. *et al.* A first-generation haplotype map of maize. *Science* **326**, 1115–1117 (2009).
8. McMullen, M.D. *et al.* Genetic properties of the maize nested association mapping population. *Science* **325**, 737–740 (2009).
9. Buckler, E.S. *et al.* The genetic architecture of maize flowering time. *Science* **325**, 714–718 (2009).
10. Laurie, C.C. *et al.* The genetic architecture of response to long-term artificial selection for oil concentration in the maize kernel. *Genetics* **168**, 2141–2155 (2004).
11. Flint, J. & Mackay, T.F. Genetic architecture of quantitative traits in mice, flies, and humans. *Genome Res.* **19**, 723–733 (2009).
12. Koornneef, M., Alonso-Blanco, C. & Vreugdenhil, D. Naturally occurring genetic variation in *Arabidopsis thaliana*. *Annu. Rev. Plant Biol.* **55**, 141–172 (2004).
13. Takahashi, Y., Teshima, K.M., Yokoi, S., Innan, H. & Shimamoto, K. Variations in Hd1 proteins, Hd3a promoters, and Ehd1 expression levels contribute to diversity of flowering time in cultivated rice. *Proc. Natl. Acad. Sci. USA* **106**, 4555–4560 (2009).
14. Turner, A., Beales, J., Faure, S., Dunford, R.P. & Laurie, D.A. The pseudo-response regulator Ppd-H1 provides adaptation to photoperiod in barley. *Science* **310**, 1031–1034 (2005).
15. Rice, W.R. Analyzing tables of statistical tests. *Evolution* **43**, 223–225 (1989).
16. Chopra, S., Athma, P., Li, X.G. & Peterson, T. A maize Myb homolog is encoded by a multicopy gene complex. *Mol. Gen. Genet.* **260**, 372–380 (1998).
17. Palaisa, K.A., Morgante, M., Williams, M. & Rafalski, A. Contrasting effects of selection on sequence diversity and linkage disequilibrium at two phytoene synthase loci. *Plant Cell* **15**, 1795–1806 (2003).
18. Valdar, W. *et al.* Genome-wide genetic association of complex traits in heterogeneous stock mice. *Nat. Genet.* **38**, 879–887 (2006).
19. Huang, G.J. *et al.* High resolution mapping of expression QTLs in heterogeneous stock mice in multiple tissues. *Genome Res.* **19**, 1133–1140 (2009).
20. Valdar, W., Holmes, C.C., Mott, R. & Flint, J. Mapping in structured populations by resample model averaging. *Genetics* **182**, 1263–1277 (2009).
21. Dickson, S.P., Wang, K., Krantz, I., Hakonarson, H. & Goldstein, D.B. Rare variants create synthetic genome-wide associations. *PLoS Biol.* **8**, e1000294 (2010).
22. Moreno, M.A., Harper, L.C., Krueger, R.W., Dellaporta, S.L. & Freeling, M. liguleless1 encodes a nuclear-localized protein required for induction of ligules and auricles during maize leaf organogenesis. *Genes Dev.* **11**, 616–628 (1997).
23. Walsh, J., Waters, C.A. & Freeling, M. 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 (1998).
24. Clark, R.M., Wagler, T.N., Quijada, P. & Doebley, J. A distant upstream enhancer at the maize domestication gene tb1 has pleiotropic effects on plant and inflorescent architecture. *Nat. Genet.* **38**, 594–597 (2006).
25. Salvi, S. *et al.* Conserved noncoding genomic sequences associated with a flowering-time quantitative trait locus in maize. *Proc. Natl. Acad. Sci. USA* **104**, 11376–11381 (2007).
26. Lettice, L.A. *et al.* A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. *Hum. Mol. Genet.* **12**, 1725–1735 (2003).
27. Flint-Garcia, S.A. *et al.* Maize association population: a high-resolution platform for quantitative trait locus dissection. *Plant J.* **44**, 1054–1064 (2005).
28. Schork, N.J., Murray, S.S., Frazer, K.A. & Topol, E.J. Common vs. rare allele hypotheses for complex diseases. *Curr. Opin. Genet. Dev.* **19**, 212–219 (2009).
29. Lai, J. *et al.* Genome-wide patterns of genetic variation among elite maize inbred lines. *Nat. Genet.* **42**, 1027–1030 (2010).
30. Swanson-Wagner, R.A. *et al.* Pervasive gene content variation and copy number variation in maize and its undomesticated progenitor. *Genome Res.* **20**, 1689–1699 (2010).

ONLINE METHODS

Materials and data. Details about how the NAM population was created and genotyped have been previously described[8,9]. We also included the public maize intermated B73-by-Mo17 (IBM) population[31] in our analysis, which resulted in 4,892 lines from 26 families. Three leaf traits (leaf length, leaf width and upper leaf angle) were scored in nine trials during the summer and winter of 2006 and 2007 in six locations: Aurora, New York, Urbana, Illinois, and Clayton, North Carolina, USA, in 2006 and 2007; Columbia, Missouri and Homestead, Florida, USA and Ponce, Puerto Rico in 2006. Leaf length was measured as the distance from the base to the tip of the leaf below the primary ear at or near flowering time. Leaf width was measured as the width of the widest section of the leaf below the primary ear at or near flowering time. Upper leaf angle was measured as the angle between the horizontal and the midrib of the first leaf below the flag leaf. Using this definition, more upright leaves have a higher leaf angle. We used the angle of the first leaf below the flag leaf as the representative of upper leaf angle for a plant because the angles of all leaves between the flag leaf and ear leaf are highly correlated[32]. The best linear unbiased predictors (BLUPs) for each line were calculated with ASREML version 2.0 software[33]. BLUPs for each line across environments were used for the overall analysis. Separate BLUPs for each environment were used to test QTLxE interaction. Leaf length and leaf width were normally distributed. The Box-Cox transformation family[34] was used to transform upper leaf angle to make the values approach normality. The subsequent analyses for leaf angle were carried out on the transformed data.

Joint linkage analysis. The method for joint linkage mapping has been previously described[9]. A brief description of joint linkage analysis used in this study is in the **Supplementary Note**. To facilitate GWAS, the method used to impute missing genotypes of NAM markers was modified. Instead of imputing missing NAM marker genotypes as the simple average of the genotype values of flanking markers[9], the average was weighted by the relative genetic distance of the missing marker from the flanking markers in this study. If the B73 genotype is coded as 0 and the non-B73 genotype is coded as 1, this result can be interpreted as the probability that a SNP comes from the non-B73 parent.

Genetic overlap between traits. The NAM design provides a unique opportunity to test the genetic overlap between different complex traits. If two traits share a QTL, the allele effects at that locus should be correlated. If the trait QTLs are different but colocalize, then the effects will be uncorrelated. Because there is no broadly accepted method for determining the confidence interval of QTL in multicross designs, we proposed a new method to calculate support interval of QTL (**Supplementary Note**). We used this method to calculate a support interval ($P < 0.05$) for each joint QTL (**Supplementary Tables 1–3**). QTLs were defined as colocalized if their support intervals ($P < 0.05$) overlapped. The correlations of effect estimates for each pair of traits at colocalized QTLs were tested. Those QTLs with a significant Pearson correlation coefficient ($P < 0.05$) are probably shared loci for different traits.

HapMap SNP imputation and projection. SNPs (1.6 million) from the maize HapMap project discovered on the NAM parents[7] were used for GWAS. Because 23% of the genotypes were missing in the set of 1.6 million SNPs, we used fastPHASE version 1.3 (ref. 35), a haplotype clustering algorithm, to impute the missing genotypes (**Supplementary Note**).

SNPs from the NAM parents were imputed onto the 4,892 RILs by first determining the physical position of the NAM markers by blasting them against a B73 reference genome (B73 RefGen_v1)[36]. For each SNP, its values for a RIL were assigned based on the SNP value of the RIL parents and on the genotype of the flanking NAM markers in that RIL. If the non-B73 parent carried the B73 allele, then the SNP was assigned a value of 0. If the non-B73 parent carried the non-B73 allele, then the SNP was assigned a value equal to the average of the flanking NAM markers weighted by its relative physical distance between the markers, similar to the method used to impute missing NAM marker values.

Genome-wide association. To further dissect the joint linkage QTLs, we conducted a GWAS within 4,892 RILs that contain 1.6 million SNPs imputed from founders. In the previous stage of joint linkage mapping, we tested linkage marker effect within each population, where we assumed in the statistical model that each founder carries a specific allele at each QTL. In the GWAS stage, we first imputed HapMap SNPs onto RILs by using information from linkage markers to infer the parent origin of each intermarker segment, and we then tested imputed SNPs across populations, where we assumed in the statistical model that the founder that shares the same SNP allele also share the same allele at the QTL. Testing imputed SNPs across populations amounts to testing whether a particular SNP is associated with phenotype independently of population (after controlling the background population difference). Therefore, by exploiting both recent and historical recombination captured by linkage markers and HapMap SNPs, respectively, we narrowed down QTLs to candidate SNPs. Thus, the goal of GWAS is to dissect QTLs into QTNs.

In this study, we conducted GWAS on the top of joint linkage analysis through a statistical procedure described below. For each chromosome and each trait, we calculated residuals from the full joint linkage model with the population term and QTL from that chromosome removed. In this way, ~90% of genetic variance was controlled in each chromosome association scan. In the stepwise regression model for joint linkage mapping, the population term was included in the model to account for the difference among populations, resulting in more reasonable markers within population estimates. We left the population term out of the model in the intermediate step of calculating a residual, but we included the population term in the final step of chromosome association scan to control the possible spurious associations caused by population differences. We then constructed a robust subsampling-based multiple SNP model to carry out association mapping, a strategy successfully applied in GWAS in mouse[18,19] and confirmed to be more robust than single-locus scanning[20]. To form each subsample, 150 lines were randomly chosen without replacement from each population. Forward regression was used to fit SNPs to the residual of each chromosome for each trait using a chromosome and trait-specific significance threshold, determined from permutations, as the stopping rule. This procedure was carried out for 100 subsamples for each chromosome and trait. By subsampling, the order of SNPs entering the model varies, compensating for the instability of model selection using forward regression and providing a measure of the reliability of position estimates. The BPP, defined as the proportion of times the SNP is included in the multiple SNP model, was used to measure the robustness of the support for SNP associations. The median of the effect size and P value of each associated SNP across all models containing that SNP was used to represent the effect size and P value of the associated SNP. The analysis was carried out using code written in Java, which is available upon request. A permutation test was used to verify that the significance levels used controlled false positives in the bootstrap analysis. Because of the time required to carry out a single bootstrap permutation, only two permutations were carried out for each chromosome. The resulting 20 permutations were pooled to generate a rough estimate of expected number of false discoveries. Using BPP ≥ 0.05, the cutoffs for leaf length and width were sufficient to keep the number of false discoveries <10 for both permutations or an average of 0.5 per chromosome per permutation. However, for leaf angle the number of false discoveries was excessive. Therefore, the P-value cutoff used for leaf angle was adjusted downward to achieve the same level observed in leaf length and width.

Overlap between QTL and GWAS. To evaluate the overlap between joint linkage QTL support regions and GWAS results, the number of associations within each QTL support interval under different BPP thresholds (BPP = 0.05, 0.1, 0.2, 0.3, 0.4 and 0.5) was calculated. Then, the proportion that would be expected by chance under each BPP cutoff was obtained by calculating the proportion of 1.6 million SNPs within QTL support intervals. Under each BPP cutoff, the binomial distribution was used to test the null hypothesis that both proportions (the actual proportion of associations within QTL support interval and the proportion expected by chance) come from the same distribution. Rejecting the null hypothesis was taken as an evidence of significant overlap between joint linkage QTL mapping and GWAS results.

Genotyping of SNPs identified from GWAS in association panel. To further examine the associations with *lg1*, *lg2* and *lg4*, seven of significant SNPs identified from GWAS were genotyped in an association panel of 282 diverse maize

inbred lines[27] using the KASPar system. Primers were designed for KASPar genotyping using a tool provided by KBiosciences (see URLs) based on the SNP locus sequence (**Supplementary Table 13**). To evaluate the ability of an assay to produce clearly distinguished clusters, artificial heterozygotes were created by pooling DNA from NAM parents according to the genotype at these HapMap SNPs (**Supplementary Table 14**). The detailed protocol for KASPAR genotyping is in the protocol provided by KBiosciences.

31. Lee, M. *et al.* Expanding the genetic map of maize with the intermated B73 x Mo17 (IBM) population. *Plant Mol. Biol.* **48**, 453–461 (2002).
32. Mickelson, S.M., Stuber, C.S., Senior, L. & Kaeppler, S.M. Quantitative trait loci controlling leaf and tassel traits in a B73 x Mo17 population of maize. *Crop Sci.* **42**, 1902–1909 (2002).
33. Gilmour, A.R., Gogel, B.J., Cullis, B.R. & Thompson, R. *ASReml User Guide: Release 2.0.* (VSN International Ltd, Hemel Hempstead, UK, 2006).
34. Box, G.E.P. & Cox, D.R. An analysis of transformations. *J. R. Stat. Soc., B* **26**, 211–252 (1964).
35. Scheet, P. & Stephens, M. 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 (2006).
36. Schnable, P.S. *et al.* The B73 maize genome: complexity, diversity, and dynamics. *Science* **326**, 1112–1115 (2009).

doi:10.1038/ng.746