DOI: 10.1534/genetics.108.090134

The Genetic Architecture of Complex Traits in Teosinte (*Zea mays* ssp. *parviglumis*): New Evidence From Association Mapping

Allison L. Weber,*[,1] William H. Briggs,*[,2] Jesse Rucker,* Baltazar M. Baltazar,[†,3] José de Jesús Sánchez-Gonzalez,[‡] Ping Feng,[§] Edward S. Buckler[,††] and John Doebley***

*Laboratory of Genetics, University of Wisconsin, Madison, Wisconsin 53706, [†]Pioneer Hi-Bred International, Tapachula, Nayarit, Mexico CP63733, [‡]Centro Universitario de Ciencias Biológicas y Agropecuarias, Universidad de Guadalajara, Jalisco, Mexico CP45110, [§]Monsanto Company, Ankeny, Iowa 50021, [††]Department of Plant Breeding and Genetics and Institute for Genomic Diversity, Cornell University, Ithaca, New York 14853 and **United States Department of Agriculture–Agricultural Research Service, Ithaca, New York 14853

Manuscript received April 9, 2008
Accepted for publication August 21, 2008

ABSTRACT

Previous association analyses showed that variation at major regulatory genes contributes to standing variation for complex traits in Balsas teosinte, the progenitor of maize. This study expands our previous association mapping effort in teosinte by testing 123 markers in 52 candidate genes for association with 31 traits in a population of 817 individuals. Thirty-three significant associations for markers from 15 candidate genes and 10 traits survive correction for multiple testing. Our analyses suggest several new putative causative relationships between specific genes and trait variation in teosinte. For example, two *ramosa* genes (*ra1* and *ra2*) associate with ear structure, and the MADS-box gene, *zagl1*, associates with ear shattering. Since *zagl1* was previously shown to be a target of selection during maize domestication, we suggest that this gene was under selection for its effect on the loss of ear shattering, a key domestication trait. All observed effects were relatively small in terms of the percentage of phenotypic variation explained (<10%). We also detected several epistatic interactions between markers in the same gene that associate with the same trait. Candidate-gene-based association mapping appears to be a promising method for investigating the inheritance of complex traits in teosinte.

THROUGH the characterization of major loss-of-function mutants, geneticists have determined the function of a vast number of genes. Despite a general knowledge of how these genes control developmental and physiological processes, very little is known about how (or if) they contribute to natural variation for complex traits. Association mapping with its high mapping resolution, its potential to sample multiple alleles, and its use of preexisting populations provides a powerful tool to investigate the role of these genes in the genetic architecture of complex traits (Risch and Merikangas 1996; Gupta *et al.* 2005; Yu and Buckler 2006). For example, association mapping in humans has found that *OCA2*, a gene responsible for ocular albinism, contributes to variation in hair and eye color (Duffy *et al.* 2007; Sulem *et al.* 2007). In several association mapping studies in plants, genes originally characterized through mutants have been found to associate with variation in complex traits such as flowering time (Thornsberry *et al.* 2001), starch pasting properties (Wilson *et al.* 2004), and vernalization response (Balasubramanian *et al.* 2006). These results suggest that genes previously characterized through mutant phenotypes might serve as good candidates in candidate-gene-based association mapping.

Previously, we detected significant associations between polymorphisms in nine candidate genes and phenotypic variation in the maize ancestor, Balsas teosinte (*Zea mays* ssp. *parviglumis*) (Weber *et al.* 2007). Our study builds upon our prior analyses in several ways, including an increase in the numbers of individuals, candidate genes, and traits. We also selected our association mapping panel to decrease the amount of population structure as compared to our prior study. With this strategy, we detected 33 associations between complex traits in teosinte and our candidate genes that survive a correction for multiple testing. These include associations between *indeterminate spikelet1* and inflorescence branching, *ramosa1* and *ramosa2* and ear structure, *sugary1* and seed oil content, and *terminal ear1* and ear length. We also observed an association between *zea agamous-like1* (*zagl1*) and ear shattering. Since *zagl1* was a target of selection during domestication, we propose that it was selected for its role in ear disarticulation. Several epistatic

[1]*Corresponding author:* Department of Genetics, North Carolina State University, Gardiner Hall 3510, Box 7614, NCSU Campus, Raleigh, NC 27695. E-mail: alweber@ncsu.edu

[2]*Present address:* Syngenta Seeds, B.V., Westeinde 62, 1601 BK Enkhuizen, The Netherlands.

[3]*Present address:* Monsanto Company, 800 N. Lindbergh Blvd., St. Louis, MO 63167.

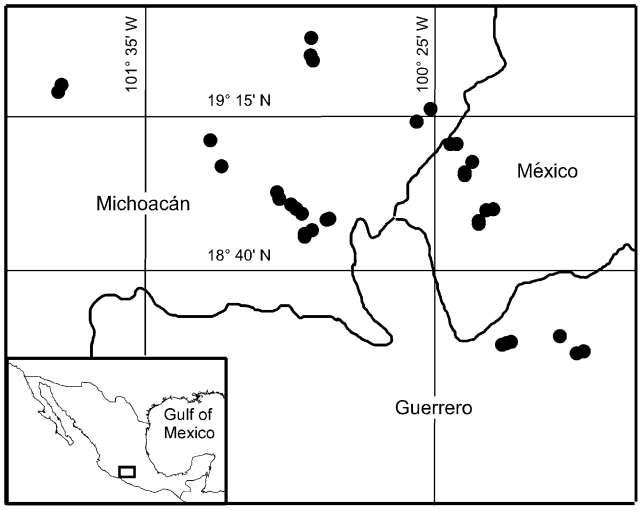

Figure 1.—Map showing region of Mexico where the local populations of Balsas teosinte were sampled. Each dot represents the location of one of the 34 local populations from which the seed for the 817 plants was collected.

interactions were detected between markers in the same candidate gene that associate with the same trait.

MATERIALS AND METHODS

Teosinte sample: A sample of 817 plants of Balsas teosinte was grown in 2004–2005 at the Pioneer Hi-Bred research station located in Tapachula, Nayarit, Mexico. The seed for these plants came from 34 local populations (a local population being a group of plants within a few hundred meters of each other and potentially interbreeding) that were found throughout the central Balsas river drainage (Figure 1; supplemental Table 1). Our goal was to sample 1020 plants in total (30 from each local population); however, 203 plants with substantial missing data were dropped from the analysis, giving a total of 817. As compared to the teosinte sample in our prior study (Weber *et al.* 2007), the current sample comes from a more restricted geographic area. We sampled a smaller area in an effort to reduce the amount of population structure. The plants were planted in a completely randomized design. None of the individuals analyzed in this study were included in our previous analysis (Weber *et al.* 2007).

Phenotypes: Thirty-one phenotypes were measured that can be grouped into five categories: flowering time (4 traits), plant architecture (5 traits), inflorescence architecture (14 traits), kernel composition (4 traits), and vegetative morphology (4 traits). Table 1 lists the traits that significantly associate with a marker after correction for multiple testing, as well as the trait fruitcase length (FCLN), which is discussed in the text. The remaining 20 traits are listed in supplemental Table 2. All lateral branch traits were measured on the second lateral branch from the top of the plant.

Genotyping: A set of 355 random genes was picked from ~10,000 low-copy-number maize ESTs without consideration as to gene function or gene type (Gardiner *et al.* 2004). These genes were sequenced using a discovery panel that consisted of 14 maize inbred lines and 16 teosinte partial inbreds (Wright *et al.* 2005). A set of 498 SNPs was selected from sequence alignments for the random genes (supplemental Table 3) and was used to control for population structure in the association analyses. A majority of these control SNPs (316) were also used for population structure analysis in our previous study. The criteria for selecting the additional 182 control SNPs followed standard procedures (Weber *et al.* 2007). A set of 52 candidate genes was selected because they have possible effects on the phenotypes under study given their known mutant phenotype in maize or other plants (Table 2; supplemental Table 4). We used sets of previously published sequence alignments for these genes to select SNPs (http://www.panzea.org). Because

TABLE 1

List of traits that were found to associate with a candidate marker

Trait	Description[a]	Units
Fruitcase length (FCLN)	Length of the female and hermaphroditic portions of the basal-most ear on the lateral branch divided by the number of fruitcases in those portions	mm
Female ear length (FERL)	Length of the female and hermaphroditic portions of the basal-most ear on the lateral branch	mm
Lateral branch internode number (LBIN)	Number of internodes that compose the lateral branch	Count
Leaf number (LFNM)	Number of leaves on the main stalk with the first leaf above ground being counted as leaf one	Count
Lateral inflorescence branch number (LIBN)	Number of branches in the tassel or inflorescence not including the central spike that terminates the lateral branch	Count
Number of fruitcases (NMFC)	Number of fruitcases in the basal-most ear of the lateral branch; no fruitcases present in branches of the ear were included in the count	Count
Oil content (OLCT)	Percentage of oil per gram of seed	%
Percentage of nondisarticulating fruitcases (NDFC)	Percentage of fruitcases that did not fully disarticulate; this trait was measured on bulk seed harvested from the mature plant (Figure 6B)	%
Tassel branch number (TBN)	Number of branches on the main tassel	Count
Tiller number (TILL)	Number of tillers at time of pollen shed	Count
Percentage of yoked fruitcases (YKFC)	Percentage of fruitcases that are yoked; this trait was measured on bulk seed harvested from the mature plant. Fruitcases are ordinarily arranged in an array, one on top of the other. Yoked fruitcases are positioned side by side (Figure 6C)	%[b]

[a] All lateral branch traits were measured on the second lateral branch.
[b] A square-root transformation was performed on the trait values.

TABLE 2

List of candidate genes that were found to associate with a trait

Gene	Gene symbol	Description
barren stalk1	*ba1*	A transcription factor that affects plant and inflorescence architecture (Gallavotti *et al.* 2004)
elongated mesocotyl1	*elm1*	A phytochromoblin synthase that affects flowering time in maize (Sawers *et al.* 2002)
indeterminate spikelet1	*ids1*	A transcription factor that affects inflorescence architecture in maize (Chuck *et al.* 1998)
ramosa1	*ra1*	A transcription factor that affects inflorescence architecture in maize (Vollbrecht *et al.* 2005)
ramosa2	*ra2*	A transcription factor that affects inflorescence architecture in maize (Bortiri *et al.* 2006)
sugary1	*su1*	An isoamylase that is involved in the biosynthesis of starch in maize (James *et al.* 1995)
teosinte branched1	*tb1*	A transcription factor that affects branching and inflorescence architecture in maize (Doebley *et al.* 1997)
terminal ear1	*te1*	An RNA-binding gene known to affect inflorescence sex and plant architecture in maize (Veit *et al.* 1998)
thick tassel dwarf1	*td1*	A leucine-rich repeat receptor-like kinase that affects plant and inflorescence architecture in maize (Bommert *et al.* 2005b)
zea agamous-like1	*zagl1*	A MADS-box transcription factor homologous to *SUPPRESSOR OF OVEREXPRESSION OF CONSTANS1* that affects flowering time in Arabidopsis (Samach *et al.* 2000)
zea apetala homolog1	*zap1*	A MADS-box transcription factor homologous to the floral homeotic gene *APETALA1* that affects inflorescence architecture in Arabidopsis (Mandel *et al.* 1992)
zea floricaula leafy1	*zfl1*	A transcription factor homologous to *LEAFY* that affects floral development and flowering time in Arabidopsis (Weigel *et al.* 1992)
zea floricaula leafy2	*zfl2*	A transcription factor homologous to *LEAFY* that affects floral development and flowering time in Arabidopsis (Weigel *et al.* 1992)
zea mays circadian1	*ZmCIR1*	A MYB repeat protein that is homologous to *CIRCADIAN 1* that affects the circadian clock and flowering time in Arabidopsis (Zhang *et al.* 2007)
zea mays gigantea	*ZmGI*	A gene of unknown function homologous to *GIGANTEA* that affects the circadian clock and flowering time in Arabidopsis (Fowler *et al.* 1999)

these candidate gene alignments come from different sources, the accessions represented in the discovery panels were variable. Similar criteria to those used for control SNP selection were used in the selection of candidate gene SNPs. A total of 123 SNP markers were developed in the 52 candidate genes. DNA extractions were accomplished using standard procedures (Briggs *et al.* 2007). SNP genotyping was performed at Genaissance Pharmaceuticals using the Sequenome MassARRAY system (Jurinke *et al.* 2002). Sequence alignments and marker context sequences are available at http://www.panzea.org.

Population structure: Population structure within our sample of 817 plants was evaluated using several statistics calculated with PowerMarker (Liu and Muse 2005). First, deviations from Hardy–Weinberg expectations for the control set of SNPs were assessed using Fisher's exact test. Second, F_{ST} was used to measure the extent of differentiation among the 34 local populations. Confidence intervals for F_{ST} were generated using 10,000 bootstrap resamplings over loci. Third, F_{IS} was calculated as a measure of recent coancestry among individuals within local populations. Again, 10,000 bootstrap resamplings over loci were used to generate confidence intervals. Fourth, we assessed the degree of correlation between geographic and genetic distance since population structure resulting from isolation-by-distance would produce such a correlation. Great circle distances between individuals using latitude and longitude were calculated using the Fields module (Nychka 2007) in the statistical computing language R (R Development Core Team 2005). The correlation coefficient between the geographic and the genetic (negative log of the proportion of shared alleles) distances was computed and its significance evaluated with the Mantel test (10,000 permutations).

Principal component analysis and a kinship matrix were computed to control for population structure and recent coancestry, respectively (Zhao *et al.* 2007). Principal component analysis was conducted with the random markers using the program EIGENSTRAT (Price *et al.* 2006). We eliminated 44 of the 498 control SNPs because they were in high LD ($r^2 > 0.5$, as defined by Hill and Robertson 1968) with another control SNP. The r^2 values were calculated using PowerMarker. The remaining set of 454 control SNPs was used for principal-component analysis. We incorporated 10 principal components in our model to describe population structure. To correct for recent coancestry or familial relatedness, a kinship matrix composed of the proportion of shared alleles for all pairwise combinations of the 817 plants was generated. These values were calculated with PowerMarker using the full set of 498 SNPs.

Testing of marker–trait associations: A mixed linear model was used to test marker–trait associations,

$$y = P\nu + S\alpha + Iu + e,$$

where y is a vector of phenotypic values, ν is a vector of fixed effects regarding population structure, α is the fixed effect for

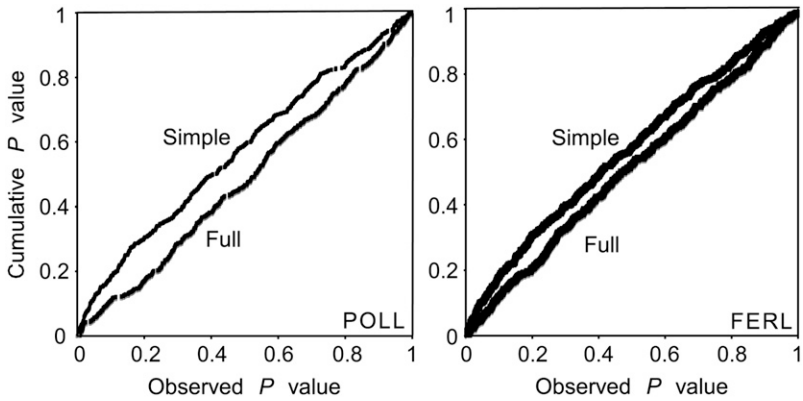

FIGURE 2.—Plots of the cumulative distributions of *P*-values for the associations between our 498 control SNPs and days to pollen shed (POLL) and female ear length (FERL). The line labeled "Simple" within each plot represents the cumulative distribution of *P*-values under a simple model that does not incorporate population structure or recent coancestry. The diagonal line represents the cumulative distribution of *P*-values under the full model that includes 10 principal components that account for population structure and a kinship matrix that accounts for recent coancestry. The *P*-value distribution for the full model follows the expected distribution under the null hypothesis of independence between the SNPs and the traits.

the candidate marker, u is a vector of the random effects pertaining to recent coancestry, and e is a vector of residuals. P is a matrix of the 10 significant principal component vectors. S is the vector of genotypes at the candidate marker, and I is an identity matrix. The variances of the random effects are assumed to be $\mathrm{Var}(u) = 2KV_g$ and $\mathrm{Var}(e) = IV_R$, where K is the kinship matrix consisting of the proportion of shared allele values, I is an identity matrix, V_g the genetic variance, and V_R the residual variance. For markers that were significantly associated with a trait, a general linear model with all of the fixed-effect terms described above was used to estimate the amount of phenotypic variation explained by each of the candidate markers, as measured by R^2. The standardized effect of each marker was also calculated by dividing the difference between the two homozygous classes by the phenotypic standard deviation of that trait. If two markers associated with the same trait, the above model was expanded to test for epistasis by including two $S\alpha$ terms for the two markers, as well as an interaction term to test for epistasis between the two markers.

This mixed linear model was used to test 1407 of the possible 3813 (123 markers × 31 traits) marker–trait pairs. Rather than testing all possible marker–trait pairs, prior knowledge regarding inferred function of the candidate gene or mutant phenotype was used to determine which traits should be tested with a given candidate gene (supplemental Table 5). For each marker–trait association, the mixed linear model described above was run in SAS using PROC MIXED (SAS INSTITUTE 1999). To assess the significance of the marker effect of each marker–trait pair, we used the F test with the denominator degrees of freedom determined by the Satterthwaite method. Residual plots were examined to determine if there were any patterns indicating that a transformation was necessary. For 29 of the 31 traits no transformation was necessary (data not shown). A square-root transformation was performed on the values for both percentage of paired spikelets (PASP) and percentage of yoked fruitcases (YKFC). The false discovery rate was used to correct for multiple testing with the exception of the vegetative morphology traits where the Bonferroni correction was used due to the small number of markers tested (STOREY 2002; STOREY and TIBSHIRANI 2003). LD among candidate markers in the same gene was assessed with r^2 values generated by PowerMarker (LIU and MUSE 2005).

The mixed linear model as described above was also used to test for associations between four traits (female ear length, FERL; leaf number, LFNM; nondisarticulating fruitcases, NDFC; and tassel branch number, TBN) and a common maize haplotype in the candidate gene *zagl1*, which is defined by five markers [PZD00020.3 (G), PZD00020.4 (T), PZD00020.2 (C), PZD00021.5 (A), and PZD00021.2 (T)]. Genotypic data were phased using PowerMarker. Individuals were coded as 0, 1, or 2 to describe how many copies of the haplotype of interest they possessed.

All of our data files (genotypes, phenotypes, seed source information, principal components, and the kinship matrix) are available on http://www.panzea.org.

RESULTS

Population structure: Several measures of population structure indicate that our sample of 817 plants is not a single unstructured Hardy–Weinberg population. First, the genotype frequencies deviate significantly from Hardy–Weinberg equilibrium ($P < 0.05$; data not shown) at ~87% (432 of 498) of the SNPs that serve as controls for population structure. Second, population differentiation (F_{ST}) among the 34 local populations is high ($F_{ST} = 0.1547 \pm 0.0108$). Third, F_{IS} is also high ($F_{IS} = 0.0805 \pm 0.0107$), suggesting recent coancestry among some of the individuals sampled. Finally, the correlation between genetic and geographical distance is significant in our study ($r = 0.3969$, $P < 0.0001$), suggesting that genetic variation is geographically structured in our sample.

Since we detected a considerable amount of population structure within our sample, we assessed the ability of a mixed linear model to control for population structure and thereby decrease the number of false positive associations (YU *et al.* 2006). This model included 10 principal components that summarize variation at 498 control SNPs. To control for recent coancestry, the model also included a kinship matrix in which the individual elements were the proportion of shared alleles at the 498 control SNPs. To assess the number of false positive associations found with both our mixed linear model and a simple model, not incorporating any control for population structure or recent coancestry, we plotted the ranked raw *P*-values by the cumulative *P*-values for associations between our 498 control SNPs and days to pollen shed (POLL) and FERL (Figure 2). There is an excess of small *P*-values with the simple model as indicated by the slight curvature of the simple line in Figure 2 for both POLL and female ear length (FERL). Both traits had approximately twice the number of significant associations (*P*-value ≤0.05) with random markers than would be expected under the null hypothesis of independence

TABLE 3

List of significant marker–trait pairs after correction for multiple testing

Trait	Gene	Marker	N^b	R^2	$2a/\sigma_P{}^c$	d/a	$2a$	d	P	FDR Q-value[d]
FERL	*zagl1*	zagl1.1	487	0.020	0.176	−2.56	1.90	−2.43	0.0054	0.0783
FERL	*ra1*	PZD00073.5	474	0.019	0.833	−1.01	9.03	−4.56	0.0062	0.0783
FERL	*ra1*	PZD00073.8	457	0.018	0.932	−0.95	10.2	−4.80	0.0068	0.0783
FERL	*te1*	PZD00006.1	465	0.019	0.300	−0.11	3.25	−0.17	0.0119	0.0984
FERL	*zap1*	PZD00022.3	455	0.019	0.435	0.87	4.71	2.06	0.0151	0.0984
FERL	*te1*	te1.3	489	0.015	0.421	1.01	4.54	2.29	0.0171	0.0984
LBIN	*zfl1*	PZB00055.1	748	0.015	0.156	−3.34	0.13	−0.22	0.0021	0.0313
LFNM	*zagl1*	PZD00020.3	699	0.012	0.476	−0.26	1.32	−0.17	0.0084	0.0773
LFNM	*zagl1*	PZD00020.4	731	0.011	0.423	0.22	1.18	0.13	0.0151	0.0773
LFNM	*elm1*	PZB00160.4	693	0.011	0.642	0.10	1.78	0.09	0.0178	0.0773
LFNM	*ZmCIR1*	PZB00232.5	749	0.009	0.301	−0.69	0.84	−0.29	0.0185	0.0773
LFNM	*zagl1*	zagl1.1	745	0.010	0.332	−0.28	0.93	−0.13	0.0200	0.0773
LFNM	*ZmGI*	PZB00049.2	752	0.010	0.394	0.65	1.10	0.36	0.0232	0.0773
LIBN	*tb1*	tb1.18	532	0.030	ND	0.31	3.42	0.53	0.0001	0.0055[e]
LIBN	*tb1*	tb1.19	497	0.031	ND	0.10	3.79	0.20	0.0003	0.0082[e]
LIBN	*ids1*	PZD00069.4	530	0.029	ND	0.40	5.19	1.04	0.0005	0.0091[e]
NMFC	*te1*	PZD00008.3	461	0.028	0.809	0.59	1.63	0.48	0.0012	0.0240
NMFC	*te1*	te1.3	489	0.022	0.583	1.21	1.17	0.71	0.0044	0.0260
OLCT	*su1*	su1.5	673	0.022	0.077	1.01	0.30	0.15	0.0008	0.0219[e]
OLCT	*su1*	su1.9	708	0.016	0.613	0.18	0.38	0.03	0.0038	0.0383
OLCT	*su1*	su1.12	699	0.015	0.537	0.20	0.34	0.03	0.0054	0.0383
OLCT	*su1*	su1.7	700	0.015	0.614	0.17	0.38	0.03	0.0056	0.0383
NDFC	*zagl1*	PZD00020.3	567	0.021	ND	−1.15	5.60	−3.21	0.0019	0.0310
NDFC	*zagl1*	PZD00021.2	606	0.019	ND	−0.97	2.72	−1.36	0.0031	0.0310
TBN	*ZmGI*	PZB00049.7	366	0.030	0.382	−0.86	10.6	−4.58	0.0016	0.0483
TBN	*ba1*	PZD00078.2	377	0.026	0.501	−0.17	13.7	−1.14	0.0026	0.0483
TBN	*tb1*	tb1.18	386	0.023	0.430	0.87	12.0	5.24	0.0047	0.0582
TBN	*ZmGI*	PZB00049.2	383	0.029	0.432	−0.47	12.1	−2.82	0.0070	0.0650
TBN	*zagl1*	PZD00019.1	363	0.021	0.683	−0.61	19.3	−5.87	0.0109	0.0712
TBN	*td1*	PZB01115.5	355	0.023	0.211	2.73	6.00	8.20	0.0115	0.0712
TILL	*zfl2*	zfl2.6	776	0.014	0.826	−1.34	4.41	−2.96	0.0007	0.0313[e]
YKFC[a]	*ra2*	Ra2_ORF.4	613	0.047	ND	−0.91	0.69	−0.31	7.88×10^{-7}	4.40×10^{-5}[e]
YKFC[a]	*ra2*	Ra2_promoter.3	587	0.028	ND	−1.11	0.48	−0.26	0.0002	0.000558[e]

[a] All values reported for this trait were calculated with trait values that have been transformed using a square-root transformation.
[b] Number of individuals with both trait and marker data.
[c] For traits with nonnormal distributions the $2a/\sigma_P$ statistic is misleading; therefore these traits are listed as having no data (ND).
[d] Only marker–trait associations with FDR Q-values ≤0.10 are shown.
[e] Association withstands the Bonferroni correction.

between the SNP and the phenotype (12.5%, POLL; 10.4%, FERL; supplemental Table 6). The diagonal lines on each plot in Figure 2, which indicate that the distribution of the P-values for associations between the random markers and each trait under the full model, follow the expected distribution under the null hypothesis; *i.e.*, ~5% of the associations are significant (5.6%, POLL; 5.6%, FERL). We concluded that a mixed linear model, including 10 principal components and the kinship matrix, adequately decreases the number of false positive associations.

Marker–trait associations: With the full model described above, we tested for associations between 123 markers in 52 candidate genes and 31 traits. The traits measured included those related to flowering time, plant architecture, inflorescence architecture, kernel composition, and vegetative morphology (Table 1, supplemental Table 2). Not all marker–trait pairs were tested. Instead, *a priori* knowledge of gene function and mutant phenotype were used to determine which marker–trait pairs should be tested (supplemental Table 5). In total, we tested 1407 of the possible 3813 marker–trait pairs.

Among the 1407 marker–trait pairs tested, 125 detectable ($P \leq 0.05$) associations were observed. Thirty-three of these associations were significant ($Q \leq 0.10$) after correction for multiple testing using the false discovery rate (Table 3). The 33 significant associations include 28 markers from 15 of the 52 candidate genes and 10 traits from four of the trait categories. Below, we discuss the significant associations for the four trait categories.

Flowering time: Forty-seven markers were tested for association with the four traits related to flowering time

(Table 1, supplemental Tables 2 and 5). Among these marker–trait pairs, 16 detectable associations were found. Six of these associations are significant after correction for multiple testing (Table 3). These include associations with 6 markers in four candidate genes (*zagl1*, *elm1*, *ZmCIR1*, and *ZmGI*) and a single trait, leaf number (LFNM). Two of the three *zagl1* markers, PZD00020.4 and PZD00020.3, are in linkage disequilibrium (LD) ($r^2 = 0.3170$), indicating that these associations are not independent. No epistasis was detected among any of these markers.

Inflorescence architecture: Fifty-nine markers were tested for association with 14 inflorescence architecture traits (Table 1, supplemental Tables 2 and 5). Eighty-five detectable associations were observed among the 826 marker–trait pairs. Twenty-one of these associations were significant after correction for multiple testing (Table 3). These include associations with six traits: female ear length (FERL), lateral inflorescence branch number (LIBN), number of fruitcases in the ear (NMFC), percentage of nondisarticulating fruitcases (NDFC), tassel branch number (TBN), and percentage of yoked fruitcases (YKFC).

Six markers from four candidate genes (*zagl1*, *ra1*, *te1*, and *zap1*) associate with FERL. These include two *ra1* markers, PZD00073.5 and PZD00073.8, which are in LD ($r^2 = 0.6866$) and thus not independent (Figure 3). No evidence for epistasis was observed among any of the markers associated with FERL. Three markers in two genes, *tb1* and *ids1*, associate with LIBN. The two *tb1* markers, tb1.18 and tb1.19, are in LD ($r^2 = 0.7485$) and thus not independent. None of the pairs of markers for LIBN showed evidence for epistasis. Two *te1* markers in linkage equilibrium associate with the NMFC (Figure 4). These two *te1* markers not only are significant individually, but also have a significant interaction effect as well ($P = 0.0071$), indicating epistasis or that the causative site is in LD with both of these markers. Two *zagl1* markers in linkage equilibrium associate with NDFC (Figures 5 and 6, A and B). These two markers have a significant interaction term ($P = 0.0025$), indicating epistasis or a causative site that is in LD with both markers. Two markers in LD ($r^2 = 0.3296$) in *ra2* associate with YKFC (Figure 6C). No epistasis was detected between these two markers. Six markers from five candidate genes (*ZmGI*, *ba1*, *tb1*, *zagl1*, and *td1*) associate with TBN. Two of these markers, PZB00049.7 and PZB00049.2, located in *ZmGI* are in LD ($r^2 = 0.6380$). No epistasis was detected between any of these markers.

Kernel composition: Thirty markers were tested for association with the four kernel composition traits (Table 1, supplemental Tables 2 and 5). Of the 120 marker–trait pairs tested, nine detectable associations were observed. Only the associations between markers in *su1* and oil content (OLCT) survived correction for multiple testing (Table 3). In *su1*, a marker in the putative promoter and a marker in the 3′-UTR (9.7 kb apart) are in LD ($r^2 = 0.658$), indicating that there is a high level of LD throughout the entire gene. There was no evidence for epistasis between the four SNPs in *su1*.

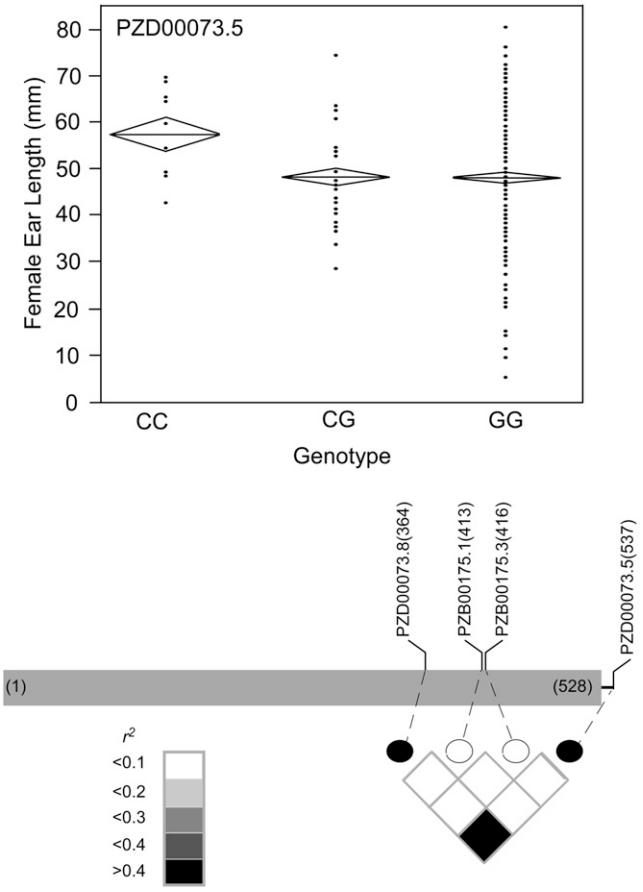

FIGURE 3.—Schematic of *ra1* showing the location in base pairs of the markers assayed and the degree of linkage disequilibrium between markers, as measured by r^2. The markers that associate with female ear length (FERL) are represented by solid circles. Shading indicates the magnitude of linkage disequilibrium. The scatterplot shows the distribution of female ear lengths for each genotypic class at marker PZD 00073.5. The mean for each genotypic class is represented by a horizontal line. The diamonds depict the standard errors of the means.

Plant architecture: Forty-nine markers were tested for association with the five plant architecture traits (Table 1, supplemental Tables 2 and 5). Of the 245 marker–trait pairs tested, 16 resulted in detectable associations. Two of these associations survive correction for multiple testing by the false discovery rate (Table 3). One marker from candidate gene *zfl1* associates with the number of internodes in the lateral branch (LBIN). One marker from candidate gene *zfl2* associates with tiller number (TILL) and survives the conservative Bonferroni correction.

***zagl1* haplotype–trait associations:** *zagl1* shows evidence of having been under selection during domestication and has a single predominant haplotype in maize inbreds that spans the entire gene (VIGOUROUX *et al.* 2002; ZHAO 2006). This haplotype is present in teosinte.

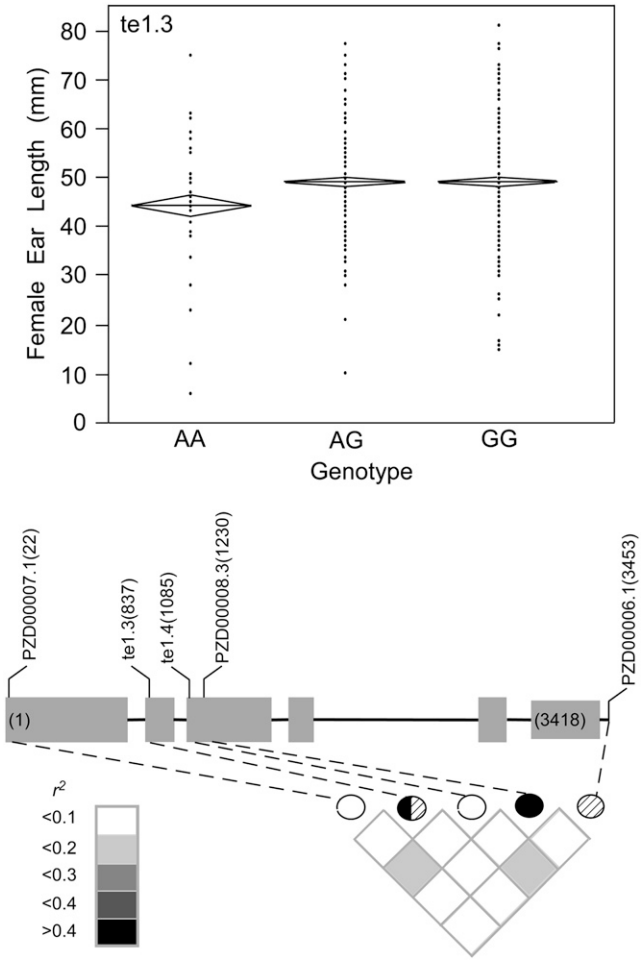

FIGURE 4.—Schematic of *te1* showing the location in base pairs of the markers assayed and the degree of linkage disequilibrium between markers, as measured by r^2. The markers that associate with female ear length (FERL) are represented by hatched circles and the markers that associate with number of fruitcases (NMFC) are represented by solid circles. Shading indicates the magnitude of the linkage disequilibrium. The scatterplot shows the distribution of female ear lengths for each genotypic class at marker te1.3. The mean for each genotypic class is represented by a horizontal line. The diamonds depict the standard errors of the means.

To evaluate if any of the four traits (FERL, LFNM, NDFC, and TBN) associate with copy number of the maize haplotype, we phased our genotypic data and tested for association. The maize haplotype was defined using five markers, PZD00020.3 (G), PZD00020.4 (T), PZD00020.2 (C), PZD00021.5 (A), and PZD00021.2 (T) (Figure 5). Individuals with two copies of the maize haplotype were designated with a two, individuals with one copy with a one, and individuals with no copies with a zero. Only NDFC was found to have a significant association with the maize haplotype (Figure 6, A and B; $P = 2.3 \times 10^{-6}$). This *P*-value was much smaller than either *P*-value observed for the two associations between the individual markers and NDFC (PZD00020.3, $P = 0.0019$; PZD00021.2, $P = 0.0031$). The mean trait values for each genotypic class indicate that the maize haplotype is recessive ($1.1 \pm 0.25\%$ for zero, $0.95 \pm 0.39\%$ for one, and $11.6 \pm 5.6\%$ for two). The maize haplotype copy number accounts for a substantial proportion of the phenotypic variance (9.4%). Selection for nondisarticulating ears during domestication could be the reason why *zagl1* has such a strong signature of selection.

Genetic associations with FERL: Six of the 33 associations detected after correction for multiple testing were between FERL and markers in four genes. To determine whether the difference in mean female ear length for the different genotypic classes at each marker was due to variation in the number of fruitcases within the ear or in fruitcase length, we examined the relationship between these markers and NMFC and FCLN. For five markers (zagl1.1, PZD00006.1, PZD00073.5, PZD00073.8, and te1.3), detectable associations were found with NMFC (Table 4), suggesting that these genes influence the number of fruitcases in the ear. For marker PZD00022.3, a detectable association was found with FCLN (Table 4), suggesting this marker influences fruitcase length.

Gene action, effects, and allele frequencies: We calculated the additive and dominance effects for each significant marker–trait association (Table 3). Most associations (48%) show partial or complete dominance ($0.50 < |d/a| < 1.25$). An additional 39% show an additive mode of inheritance with a d/a ratio between -0.5 and $+0.5$. Relatively few associations (12%) show evidence for overdominance ($|d/a| > 1.25$). We also calculated the effect sizes in terms of the percentage of the phenotypic variance that the marker explained (R^2) and the difference between the two homozygous genotypic classes in units of phenotypic standard deviations. Using R^2, all effects are small, ranging from 0.9 to 4.7%. Similarly, most effects explain only a fraction of a phenotypic standard deviation (Table 3).

Some of the traits that we measured describe morphology differences between maize and teosinte that are hypothesized to have been under selection during domestication. For each significant association regarding one of these traits, the allele that confers a more maize-like phenotype in teosinte was identified. These alleles may have been under selection during domestication and thus brought to a higher frequency in maize. To test this hypothesis, we determined the allele frequencies in our teosinte sample as well as in a sample of maize landraces (Table 5). The landrace frequency was calculated using genotypic data for 1132 maize landraces samples (noncommercial varieties found throughout the pre-Columbian range of maize in the Americas) from the Maize Diversity Database (http://www.panzea.org). Alleles associated with a maize-like phenotype are at a higher frequency in the maize landraces than in teosinte for only 6 of the 15 markers. Thus, there is no consistent trend for the allele associated with a maize-like phenotype to be at a higher frequency in maize.

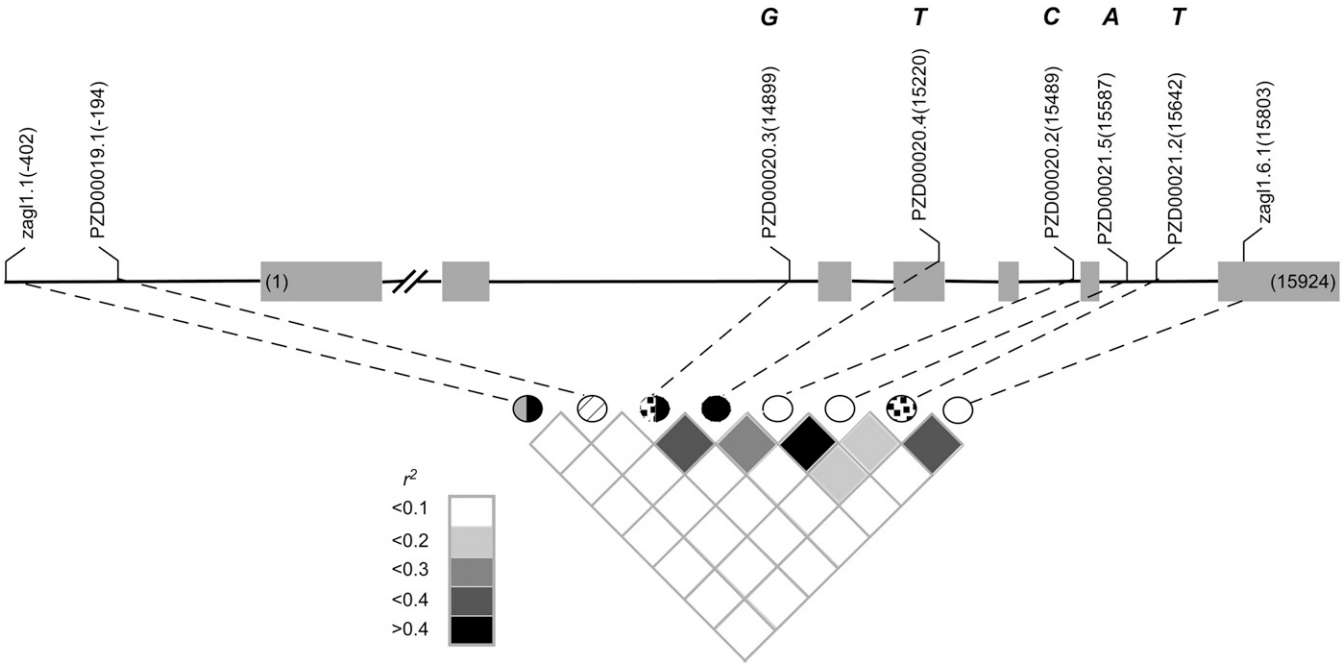

FIGURE 5.—Schematic of *zagl1* showing the location in base pairs of the markers assayed and the degree of linkage disequilibrium between markers, as measured by r^2. The marker that associates with female ear length (FERL) is represented by a partially gray-shaded circle and the marker that associates with tassel branch number (TBN) is represented by a hatched circle. The markers that associate with leaf number (LFNM) are represented by solid circles and the markers that associate with percentage of nondisarticulating fruitcases (NDFC) are represented by dotted circles. Shading indicates the magnitude of the linkage disequilibrium.

DISCUSSION

Associations and effects: The 33 significant associations identified by our study not only provide a good list of candidate genes for further investigation, but also provide some general insight into the genetic architecture of complex traits in teosinte. All of our 33 significant associations account for relatively small proportions of the phenotypic variance (Table 3). These values are similar or slightly larger than those calculated for marker–trait associations in other wild populations (DWORKIN *et al.* 2005; GONZÁLEZ-MARTÍNEZ *et al.* 2007). There are a few possible explanations for these small values. First, if the marker assayed is not the causative site but in LD with the causative site or haplotype, the R^2 value will be an underestimate of the actual effect (LAI *et al.* 1994: NIELSEN and WEIR 1999). Second, the traits may have low heritabilities such that most of the variance is environmental. Third, it is possible that these associations are actually due to alleles of small effect.

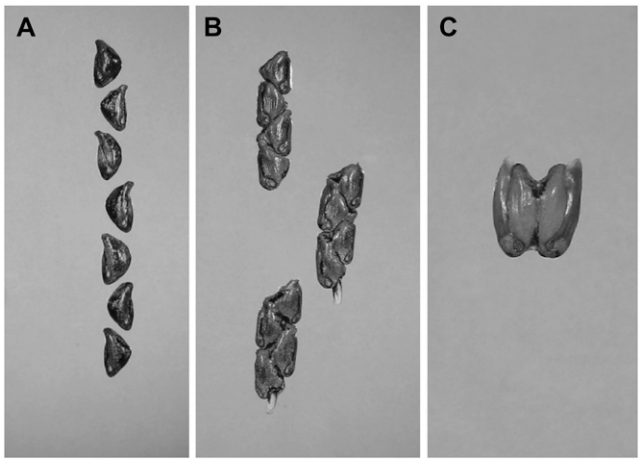

FIGURE 6.—Fruitcase phenotypes. (A) Teosinte fruitcases that disarticulate (separate) upon maturity. (B) Nondisarticulated fruitcases (fruitcases that do not separate upon maturity). (C) Yoked fruitcases (fruitcases that develop side by side as opposed to in an array on top of each other).

If one measures effects in terms of phenotypic standard deviations, then once again most effects are small, representing only a fraction of a phenotypic standard deviation (Table 3). The magnitudes of these effects are mostly similar to those calculated for other complex traits such as bristle number (LAI *et al.* 1994) and life span (DE LUCA *et al.* 2003) in Drosophila. However, a few of the effects appear somewhat larger: FERL-*ra1* ($0.932\sigma_p$), NMFC-*te1* ($0.809\sigma_p$), and TILL-*zfl2* ($0.826\sigma_p$). This observation suggests the possibility that teosinte might harbor some large effect alleles that selection could have acted upon during domestication and that a modest number of gene substitutions for alleles of such effects could change the population mean substantially over a relatively small number of generations.

Comparison of results with a previous study: The teosinte sample for this study was collected from a much smaller geographic region as compared to the sample in

TABLE 4

Associations with FERL and correlated traits

Marker	Genotype	FERL	FCLN	NMFC
zagl1.1	AA	48.5	5.23	9.17
	AC	47.0	5.28	8.91
	CC	50.4	5.40	9.30
	P-value	0.0054	0.3299	0.0283
PZD00073.5	CC	57.7	5.53	10.50
	CG	48.6	5.35	9.14
	GG	48.7	5.32	9.09
	P-value	0.0062	0.5046	0.0349
PZD00073.8	CC	58.7	5.50	10.7
	CT	48.8	5.37	9.14
	TT	48.5	5.30	9.10
	P-value	0.0068	0.7984	0.0301
PZD00006.1	CC	45.9	5.36	8.53
	CG	47.4	5.27	8.92
	GG	49.2	5.32	9.20
	P-value	0.0119	0.8844	0.0412
PZD00022.3	CC	44.9	5.12	8.73
	CT	49.3	5.33	9.21
	TT	49.6	5.37	9.20
	P-value	0.0151	0.02	0.3564
te1.3	AA	44.6	5.36	8
	AG	49.2	5.28	9.29
	GG	49.2	5.35	9.17
	P-value	0.0171	0.7968	0.0026

our previous teosinte association mapping study (Weber *et al.* 2007). A comparison of population structure statistics for this study with our previous study indicates that this change in sampling did reduce the amount of population structure. First, the percentage of control markers found to deviate from Hardy–Weinberg equilibrium was lower in this population (87%) as compared to the previous one (94%). Second, the population differentiation (F_{ST}) among the 34 local populations sampled in this study ($F_{ST} = 0.1547 \pm 0.0108$) was lower than that reported for the 74 local populations in our earlier study ($F_{ST} = 0.1690 \pm 0.005$). Third, F_{IS} was also lower in this study ($F_{IS} = 0.0805 \pm 0.0107$) than in the previous one ($F_{IS} = 0.0999 \pm 0.011$). Finally, genetic and geographical distances were not as strongly correlated in this study ($r = 0.3969$) as compared to our previous study ($r = 0.4605$).

The decrease in the amount of population structure within this sample, as well as the larger number of individuals (817 compared to 584) assayed may give this study more power to detect associations as compared to our previous study. Yet, despite less population structure as compared to our previous study, both studies identified approximately the same proportion of significant marker–trait associations (2.34%, 33 of 1407; 1.91%, 10 of 523; Weber *et al.* 2007). This result confirms that the mixed model effectively controls for structure over a range of population structuring (Yu *et al.* 2006; Zhao *et al.* 2007).

Despite the fact that both teosinte association mapping studies identified a number of significant associations that survived correction for multiple testing, there was only one repeated marker–trait association (FERL-te1.3) among the nine marker–trait pairs that were assayed in both studies. There are several reasons why so few associations were repeated. First, the phenotyping was performed by different individuals in the two studies, possibly leading to subtle differences in how the

TABLE 5

Frequency of and additive effects associated with the alleles associated with maize-like phenotypes

Markers	Allele	Trait[a]	Teosinte	Maize	α[b]
PZD00022.3	T	FERL(LEN)	0.66	0.41	−1.024
PZD00073.5	C	FERL(NUM)	0.06	0.00	0.316
PZD00073.8	C	FERL(NUM)	0.06	0.01	0.658
PZD00006.1	G	FERL(NUM)	0.90	0.19	0.165
zagl1.1	C	FERL(NUM)	0.69	0.005	0.4727
te1.3	G	FERL(NUM)/NMFC	0.80	0.93	0.258/0.055
tb1.18	T	LIBN	0.38	0.10	−0.978
tb1.19	C	LIBN	0.33	0.10	−1.241
PZD00069.4	A	LIBN	0.25	0.61	−1.452
PZD00008.3	T	NMFC	0.91	0.88	0.054
PZD00020.3	G	NDFC	0.15	0.99	0.371
PZD00021.2	T	NDFC	0.35	0.91	0.560
zfl2.6	G	TILL	0.86	0.74	−0.087
Ra2_ORF.4	A	YKFC	0.07	0.55	0.013
Ra2_promoter.3	C	YKFC	0.08	0.95	0.002

[a] Associations with female ear length correlated with differences in the number of fruitcases are designated as FERL(NUM). Associations with female ear length correlated with differences in fruitcase length are designated FERL(LEN).

[b] The additive effect was calculated as $\alpha_B = p_B G_{BB} + p_b G_{Bb} - \mu_G$, where p_i is the allele frequency of allele i, G_{ij} the phenotypic mean of the genotypic class ij, and μ_G the overall phenotypic mean.

traits were measured. Second, different sets of plants representing different portions of the Balsas teosinte range were sampled in the two studies. Third, the plants were grown in different environments (Hawaii *vs.* Mexico) in the two experiments. Finally, these two teosinte samples had different amounts of population structure for which the analyses had to correct. All of these factors could contribute to true positive associations not being detected in one study or the other. Also, while only one of the nine repeated tests was significant in both studies, three others were significant in one study and nearly significant in the other ($P < 0.1$) (supplemental Table 7), suggesting that they are also true positives.

In this study, similar to our previous study, we identified the frequency of the allele at each SNP associated with the maize-like phenotype in both our teosinte sample and a sample of maize landraces. If selection acted on these genes during domestication, one would expect this allele to be at a higher frequency in the maize landraces than in teosinte. In the previous study, all six maize-phenotype alleles were at a higher frequency in maize relative to teosinte (Weber *et al.* 2007). In this study, this was true for only 40.0% (6 of 15) of the maize-phenotype alleles. Since results from the two studies are not consistent, perhaps the phenomenon observed in the first study was merely an artifact of small sample size.

Effects on ear length: Several of our associations are with female ear length (FERL). Changes in length can be due to differences in fruitcase number (NMFC) or fruitcase length (FCLN). To determine which of these two aspects of female ear length contributed to the associations, we compared the mean trait value of several correlated phenotypes (FERL, FCLN, and NMFC) for each genotypic class at each marker that associated with FERL. Our results suggest that differences in FERL associated with variation in *ra1*, *te1*, and *zagl1* are due to changes in the number of fruitcases in the ear, while effects associated with variation in *zap1* are a function of fruitcase length. Although further work is needed to validate these trends, our results suggest that some genes affect ear length by influencing fruitcase number, while others influence ear length by influencing fruitcase length (size).

Functions for candidate genes: We detected novel associations involving markers in several genes that might not have been predicted from prior knowledge about these genes. First, markers in *ra1* associated with female ear length and specifically with the number of fruitcases in the ear. This association was unexpected since the function of *ra1* in maize is to impose short branch identity on secondary branch meristems (Vollbrecht *et al.* 2005). In *ra-R* mutants, normally determinate secondary branches in the tassel and ear are long and indeterminate. Our association would be explicable in terms of *ra1* function if teosinte ear meristems are developmentally equivalent to maize secondary branch meristems. *ra1* is an EPF zinc-finger transcription factor with both zinc-finger and EAR domains, which are hypothesized to be important for the function of *ra1* (Vollbrecht *et al.* 2005). The significant associations we observed involve a SNP located in the 3′-UTR (PZD00073.5) and a missense substitution (PZD00073.8) that changes a serine to a proline at amino acid site 122. The latter change is not within either the zinc-finger or the EAR domain. Previous studies have indicated that *ra1* was under selection during domestication (Vollbrecht *et al.* 2005). Increase in the number of fruitcases (kernels) in the ear may have been the target trait.

zagl1 is a MADS-box transcription factor, which is homologous to *SUPRESSOR OF OVEREXPRESSION OF CONSTANS 1* (*SOC1*), a promoter of flowering in Arabidopsis (Samach *et al.* 2000; Vigouroux *et al.* 2002). The function of *zagl1* in maize is unknown. Our association mapping suggests that *zagl1* has pleiotropic effects on female ear length (FERL), leaf number (LFNM), percentage of nondisarticulating fruitcases (NDFC), and tassel branch number (TBN) in teosinte (Figure 5). Since *SOC1* influences flowering time in Arabidopsis, it is not surprising that *zagl1* associates with FERL, LFNM, and TBN in teosinte. However, the association between *zagl1* and NDFC suggests an additional function for *zagl1*. Teosinte ears shatter at maturity for their seed to be dispersed, while maize ears stay intact for human harvest. We found that a maize-like haplotype (as defined by five SNPs) strongly associates with an increase in the percentage of nondisarticulating fruitcases in teosinte. The maize-like haplotype is recessive with respect to other haplotypes. Interestingly, other MADS-box genes have been implicated in disarticulation in both Arabidopsis (Ferrandiz *et al.* 2000; Liljegren *et al.* 2000) and tomato (Mao *et al.* 2000). Since *zagl1* shows evidence of selection during domestication (Vigouroux *et al.* 2002), ear shattering may have been the target trait.

Given that we observed the *zagl1* maize-like haplotype in teosinte, there is a concern that the teosinte individuals possessing this haplotype are admixed with maize, a phenomenon that could bias our results. To investigate whether the presence of the *zagl1* maize-like haplotype is due to recent admixture with maize, we analyzed the degree of admixture between our 817 teosinte plants and 277 Mexican landrace plants for the control SNPs, using a Bayesian, model-based approach (Pritchard *et al.* 2000) (data not shown). A Wilcoxon rank-sum test found no significant difference ($P = 0.4471$) in the proportion of maize ancestry between teosinte plants with *vs.* without the *zagl1* maize-like haplotype. This result suggests that the presence of the *zagl1* maize-like haplotype in teosinte is not due to a recent introgression.

su1 encodes an isoamylase involved in the biosynthesis of starch in maize (James *et al.* 1995). We detected significant associations between markers in *su1* and oil content rather than starch content. Our association be-

tween *su1* and oil content is consistent with the temporal coordination of both starch and oil metabolism during seed development in maize (Lee *et al.* 2002). *bt2*, another gene involved in starch metabolism, has also been shown to associate with oil production (Wilson *et al.* 2004). There is evidence that *su1* was under selection during domestication or subsequent maize improvement (Whitt *et al.* 2002). Our results suggest that some aspect of kernel composition may have been the target trait.

Several other associations, including LIBN-*ids1*, YKFC-*ra2*, FERL-*te1*, and NMFC-*te1*, suggest new roles for these genes in teosinte trait variation. *indeterminate spikelet1* (*ids1*) has been previously characterized as an *APETEALA2*-like transcription factor that controls floret number in both the male and the female inflorescences of maize (Chuck *et al.* 1998). Although differences in inflorescence branch number were not described by Chuck *et al.*, the expression of *ids1* in various lateral organ primordia in the inflorescence makes it plausible that *ids1* may control several levels of branching in the inflorescence. *ramosa2* (*ra2*) is also a transcription factor that has been found to affect inflorescence architecture in maize (Bortiri *et al.* 2006). The expression of *ra2* in maize inflorescence and spikelet meristems is consistent with the association we detected between this gene and the percentage of yoked fruitcases (Figure 6C), a trait that is related to spikelet formation in the ear. Finally, we repeated an association between *terminal ear1* (*te1*) and female ear length that was observed in our previous study (Weber *et al.* 2007). Moreover, in our study, we were able to attribute the variation in ear length to an increased number of fruitcases. *te1* is a RNA-binding protein that is hypothesized to control internode initiation and number in the maize stalk (Veit *et al.* 1998). Since fruitcases are modified internodes, a role for *te1* in fruitcase proliferation in the ear is reasonable given that the maize mutant alters internode number in the main stalk.

Other observed associations in our study cleanly fit the known functions of the candidate genes. An association between tassel branch number and *ba1* is expected since *ba1* mutants lack tassel branches (Gallavotti *et al.* 2004). An association between tassel branch number and *td1* is expected since *td1* mutants have extra tassel branches (Bommert *et al.* 2005a,b). Associations between leaf number and both *ZmCIR1* and *ZmGI* are expected since these genes are sequence homologs to two known flowering-time genes in Arabidopsis and flowering time and leaf number are strongly correlated traits (Fowler *et al.* 1999; Zhang *et al.* 2007). Finally, an association between *zfl2* and tillering is consistent with the observation that the rice homolog of *zfl2* (*rfl*) controls tillering in rice (Rao *et al.* 2008).

Overall, our results provide evidence that candidate-gene-based association mapping is a powerful tool for identifying genes contributing to natural variation in teosinte. Both genes for which there are characterized mutants in maize and genes known only through sequence homology to characterized mutants in other species provide useful candidates. In the case of *zagl1* and *su1*, candidate-gene-based association mapping suggests new functions for known genes. To confirm that these genes truly control the unanticipated phenotype, more detailed analyses such as QTL fine-mapping, gene-expression, and protein-function assays are required. Such additional analyses can not only confirm (or disprove) the associations but also potentially identify the specific causative polymorphism. Once validated by fine-mapping or molecular assays, our association-mapping results will help refine the understanding of the genetic architecture of complex traits in the progenitor of maize, teosinte.

We thank Bret Payseur and Brian Yandell for comments and discussion. We acknowledge Pioneer International for providing field space. We also thank the Monsanto Company for providing us access to the equipment that was used to measure the kernel composition traits. This work was funded by National Science Foundation grant DBI-0321467, National Institutes of Health grant GM-58816, and U.S. Department of Agriculture Hatch grant WIS04772.

LITERATURE CITED

Balasubramanian, S., S. Sureshkumar, M. Agrawal, T. P. Michael, C. Wessinger *et al.*, 2006 The PHYTOCHROME C photoreceptor gene mediates natural variation in flowering and growth responses of *Arabidopsis thaliana*. Nat. Genet. **38:** 711–715.

Bommert, P., N. Satoh-Nagasawa, D. Jackson and H. Hirano, 2005a Genetics and evolution of inflorescence and flower development in grasses. Plant Cell Physiol. **46:** 69–78.

Bommert, P., C. Lunde, J. Nardmann, E. Vollbrecht, M. Running *et al.*, 2005b *thick tassel dwarf1* encodes a putative maize ortholog of the *Arabidopsis CLAVATA1* leucine-rich repeat receptor-like kinase. Development **132:** 1235–1245.

Bortiri, E., G. Chuck, E. Vollbrecht, T. Rocheford, R. Martienssen *et al.*, 2006 *ramosa2* encodes a LATERAL ORGAN BOUNDARY domain protein that determines the fate of stem cells in branch meristems of maize. Plant Cell **18:** 574–585.

Briggs, W. H., M. D. McMullen, B. S. Gaut and J. D. Doebley, 2007 Advance QTL mapping in a RIL resource for positional cloning of maize domestication genes. Genetics **177:** 1915–1928.

Chuck, G., R. B. Meeley and S. Hake, 1998 The control of maize spikelet meristem fate by the *APETALA2*-like gene *indeterminate spikelet1*. Genes Dev. **12:** 1145–1154.

De Luca, M., N. V. Roshina, G. L. Geiger-Thornsberry, R. F. Lyman, E. G. Pasyukova *et al.*, 2003 Dopa decarboxylase (Ddc) affects variation in Drosophila longevity. Nat. Genet. **34:** 429–433.

Doebley, J., A. Stec and L. Hubbard, 1997 The evolution of apical dominance in maize. Nature **386:** 485–488.

Duffy, D. L., G. W. Montgomery, W. Chen, Z. Z. Zhao, L. Le *et al.*, 2007 A three single-nucleotide polymorphism haplotype in intron 1 of *OCA2* explains most human eye-color variation. Am. J. Hum. Genet. **80:** 241–252.

Dworkin, I., A. Palsson and G. Gibson, 2005 Replication of an Egfr-Wing shape association in a wild-caught cohort of *Drosophila melanogaster*. Genetics **169:** 2115–2125.

Ferrandiz, C., S. J. Liljegren and M. F. Yanofsky, 2000 Negative regulation of the *SHATTERPROOF* genes by *FRUITFULL* during Arabidopsis fruit development. Science **289:** 436–438.

Fowler, S., K. Lee, H. Onouchi, A. Samach, K. Richardson *et al.*, 1999 *GIGANTEA:* a circadian clock-controlled gene that regulates photoperiodic flowering in *Arabidopsis* and encodes a protein with several possible membrane-spanning domains. EMBO J. **18:** 4679–4688.

Gallavotti, A., Q. Zhao, J. Kyozuka, R. B. Meeley, M. K. Ritter *et al.*, 2004 The role of *barren stalk1* in the architecture of maize. Nature **432:** 630–635.

Gardiner, J., S. Schroeder, M. L. Polacco, H. Sanchez-Villeda, Z. Fang *et al.*, 2004 Anchoring 9,371 maize expressed sequence tagged unigenes to the bacterial artificial chromosome contig map by two-dimensional overgo hybridization. Plant Physiol. **134:** 1317–1326.

González-Martínez, S. C., N. C. Wheeler, E. Ersoz, C. D. Nelson and D. B. Neale, 2007 Association genetics in *Pinus taeda* L. I. Wood property traits. Genetics **175:** 399–409.

Gupta, P., S. Rustgi and P. Kulwal, 2005 Linkage disequilibrium and association studies in higher plants: present status and future prospects. Plant Mol. Biol. **57:** 461.

Hill, W. G., and A. Robertson, 1968 Linkage disequilibrium in finite populations. Theor. Appl. Genet. **38:** 226–231.

James, M. G., D. S. Robertson and A. M. Myers, 1995 Characterization of the maize gene *sugary1*, a determinant of starch composition in kernels. Plant Cell **7:** 417–429.

Jurinke, C., D. van den Boom, C. R. Cantor and H. Koster, 2002 The use of MassARRAY technology for high throughput genotyping. Adv. Biochem. Eng. Biotechnol. **77:** 58–74.

Lai, C., R. F. Lyman, A. D. Long, C. H. Langley and T. F. C. Mackay, 1994 Naturally occurring variation in bristle number and DNA polymorphism at the *scabrous* locus of *Drosophila melanogaster*. Science **266:** 1697–1702.

Lee, J., M. E. Williams, S. V. Tingey and J. A. Rafalski, 2002 DNA array profiling of gene expression changes during maize embryo development. Funct. Integr. Genomics **2:** 13–27.

Liljegren, S. J., G. S. Ditta, Y. Eshed, B. Savidge, J. L. Bowman *et al.*, 2000 *SHATTERPROOF* MADS-box genes control seed dispersal in Arabidopsis. Nature **404:** 766.

Liu, K., and S. V. Muse, 2005 PowerMarker: an integrated analysis environment for genetic marker analysis. Bioinformatics **21:** 2128–2129.

Mandel, M. A., C. Gustafson-Brown, B. Savidge and M. F. Yanofsky, 1992 Molecular characterization of the *Arabidopsis* floral homeotic gene *APETALA1*. Nature **360:** 273–277.

Mao, L., D. Begum, H.-w. Chuang, M. A. Budiman, E. J. Szymkowiak *et al.*, 2000 *JOINTLESS* is a MADS-box gene controlling tomato flower abscission zone development. Nature **406:** 910.

Nielsen, D. M., and B. S. Weir, 1999 A classical setting for associations between markers and loci affecting quantitative traits. Genet. Res. **74:** 271–277.

Nychka, D., 2007 Fields: tools for spatial data. R package version 3.5. http://www.image.ucar.edu/GSP/Software/Fields.

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

Pritchard, J. K., M. Stephens, N. A. Rosenberg and P. Donnelly, 2000 Association mapping in structured populations. Am. J. Hum. Genet. **67:** 170–181.

R Development Core Team, 2005 R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna.

Rao, N. N., K. Prasad, P. R. Kumar and U. Vijayraghavan, 2008 Distinct regulatory role for *RFL*, the rice *LFY* homolog, in determining flowering time and plant architecture. Proc. Natl. Acad. Sci. USA **105:** 3646–3651.

Risch, N., and K. Merikangas, 1996 The future of genetic studies of complex human diseases. Science **273:** 1516–1517.

Samach, A., H. Onouchi, S. E. Gold, G. S. Ditta, Z. Schwarz-Sommer *et al.*, 2000 Distinct roles of CONSTANS target genes in reproductive development of *Arabidopsis*. Science **288:** 1613–1616.

SAS Institute, 1999 *SAS/STAT User's Guide*, Version 8. SAS Institute, Cary, NC.

Sawers, R. J. H., P. J. Linley, P. R. Farmer, N. P. Hanley, D. E. Costich *et al.*, 2002 *elongated mesocotyl1*, a phytochrome-deficient mutant of maize. Plant Physiol. **130:** 155–163.

Storey, J. D., 2002 A direct approach to false discovery rates. J. R. Stat. Soc. Ser. B **64:** 479–498.

Storey, J. D., and R. Tibshirani, 2003 Statistical significance for genomewide studies. Proc. Natl. Acad. Sci. USA **100:** 9440–9445.

Sulem, P., D. F. Gudbjartsson, S. N. Stacey, A. Helgason, T. Rafnar *et al.*, 2007 Genetic determinants of hair, eye and skin pigmentation in Europeans. Nat. Genet. **39:** 1443.

Thornsberry, J. M., M. M. Goodman, J. Doebley, S. Kresovich, D. Nielson *et al.*, 2001 *Dwarf8* polymorphisms associate with variation in flowering time. Nat. Genet. **28:** 286–289.

Veit, B., S. P. Briggs, R. J. Schmidt, M. F. Yanofsky and S. Hake, 1998 Regulation of leaf initiation by the *terminal ear 1* gene of maize. Nature **393:** 166–168.

Vigouroux, Y., M. McMullen, C. T. Hittinger, K. Houchins, L. Schulz *et al.*, 2002 Identifying genes of agronomic importance in maize by screening microsatellites for evidence of selection during domestication. Proc. Natl. Acad. Sci. USA **99:** 9650–9655.

Vollbrecht, E., P. S. Springer, L. Goh, E. S. Buckler, Iv and R. Martienssen, 2005 Architecture of floral branch systems in maize and related grasses. Nature **436:** 1119–1126.

Weber, A., R. M. Clark, L. Vaughn, J. de Jesus Sanchez-Gonzalez, J. Yu *et al.*, 2007 Major regulatory genes in maize contribute to standing variation in teosinte (*Zea mays* ssp. *parviglumis*). Genetics **177:** 2349–2359.

Weigel, D., J. Alvarez, D. R. Smyth, M. F. Yanofsky and E. M. Meyerowitz, 1992 *LEAFY* controls floral meristem identity in *Arabidopsis*. Cell **69:** 843–859.

Whitt, S. R., L. M. Wilson, M. I. Tenaillon, B. S. Gaut and E. S. Buckler, IV, 2002 Genetic diversity and selection in the maize starch pathway. Proc. Natl. Acad. Sci. USA **99:** 12959–12962.

Wilson, L. M., S. R. Whitt, A. M. Ibanez, T. R. Rocheford, M. M. Goodman *et al.*, 2004 Dissection of maize kernel composition and starch production by candidate gene association. Plant Cell **16:** 2719–2733.

Wright, S. I., I. V. Bi, S. G. Schroeder, M. Yamasaki, J. F. Doebley *et al.*, 2005 The effects of artificial selection on the maize genome. Science **308:** 1310–1314.

Yu, J., and E. S. Buckler, 2006 Genetic association mapping and genome organization of maize. Curr. Opin. Biotechnol. **17:** 155.

Yu, J., G. Pressoir, W. H. Briggs, I. Vroh Bi, M. Yamasaki *et al.*, 2006 A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nat. Genet. **38:** 203–208.

Zhang, X., Y. Chen, Z.-Y. Wang, Z. Chen, H. Gu *et al.*, 2007 Constitutive expression of *CIR1 (RVE2)* affects several circadian-regulated processes and seed germination in Arabidopsis. Plant J. **51:** 512–525.

Zhao, K., M. J. Aranzana, S. Kim, C. Lister, C. Shindo *et al.*, 2007 An *Arabidopsis* example of association mapping in structured samples. PLoS Genet. **3:** e4.

Zhao, Q., 2006 Molecular population genetics of maize regulatory genes during domestication. Ph.D. Dissertation, University of Wisconsin, Madison, WI.

Communicating editor: M. Nordborg