Reshaping of the maize transcriptome by domestication

Ruth Swanson-Wagner^{a,1}, Roman Briskine^{b,1}, Robert Schaefer^c, Matthew B. Hufford^d, Jeffrey Ross-Ibarra^{d,e}, Chad L. Myers^b, Peter Tiffin^a, and Nathan M. Springer^{a,2}

^aMicrobial and Plant Genomics Institute, Department of Plant Biology, University of Minnesota, Saint Paul, MN 55108; ^bDepartment of Computer Science and Engineering, University of Minnesota, Minneapolis, MN 55455; ^cBiomedical Informatics and Computational Biology Graduate Program, University of Minnesota, Rochester, MN 55904; and ^dDepartment of Plant Sciences and ^eGenome Center, University of California, Davis, CA 95616

Edited by Hugo K. Dooner, Waksman Institute, Rutgers University, Piscataway, NJ, and approved June 12, 2012 (received for review February 2, 2012)

Through domestication, humans have substantially altered the morphology of Zea mays ssp. parviglumis (teosinte) into the currently recognizable maize. This system serves as a model for studying adaptation, genome evolution, and the genetics and evolution of complex traits. To examine how domestication has reshaped the transcriptome of maize seedlings, we used expression profiling of 18,242 genes for 38 diverse maize genotypes and 24 teosinte genotypes. We detected evidence for more than 600 genes having significantly different expression levels in maize compared with teosinte. Moreover, more than 1,100 genes showed significantly altered coexpression profiles, reflective of substantial rewiring of the transcriptome since domestication. The genes with altered expression show a significant enrichment for genes previously identified through population genetic analyses as likely targets of selection during maize domestication and improvement; 46 genes previously identified as putative targets of selection also exhibit altered expression levels and coexpression relationships. We also identified 45 genes with altered, primarily higher, expression in inbred relative to outcrossed teosinte. These genes are enriched for functions related to biotic stress and may reflect responses to the effects of inbreeding. This study not only documents alterations in the maize transcriptome following domestication, identifying several genes that may have contributed to the evolution of maize, but highlights the complementary information that can be gained by combining gene expression with population genetic analyses.

The domestication of maize from its wild progenitor is a model system for investigating domestication, genome evolution, and response to selection (1–4). Cytogenetic and molecular analyses of maize domestication have identified Zea mays ssp. parviglumis (hereafter teosinte) as the direct wild progenitor of maize and indicate these lineages diverged ~9,000 generations ago (5, 6). Despite their recent divergence, maize exhibits substantial phenotypic differences from its wild progenitor, reflecting rapid and pronounced evolutionary change (2, 5). Identification of genetic changes underlying these phenotypic differences will give insight into the genetic architecture of complex traits (3, 7), characterize response to selection (8), and provide resources for maize improvement (3).

Both quantitative trait locus (QTL) mapping and molecular population genetic scans have identified numerous genomic regions that underlie maize domestication (3, 9–14). Molecular characterization of QTL has identified genes that appear to be responsible for several of the morphological or phenological differences between maize and teosinte, including *tb1* (15), *tga1* (16, 17), *zfl2* (18), *ba1* (19), and *ra1* (20). A number of genes with putative regulatory function have been identified as potential domestication genes (3, 13), and a recent genome-wide analysis of maize and teosinte identified numerous selected regions devoid of annotated genes (14). These data are consistent with suggestions that regulation of gene expression has played an important role in the evolution of maize (3, 21–23). The importance of regulatory, as opposed to structural gene, changes is also

consistent with the broader hypothesis that regulatory differences are fundamental to the evolution of morphological and developmental diversity (24).

To investigate the evolution of gene expression that accompanied maize domestication, we examined the transcriptome of 38 diverse maize inbred lines and 24 teosinte accessions. Our specific objectives were twofold. First, we tested for significant betweentaxa differences in expression, which may occur as a result of directional selection. Second, we used coexpression network analyses to test whether domestication has rewired the transcriptional network, causing changes in the covariance of gene expression. We find evidence for significant changes in both gene expression levels and coexpression relationships following domestication and identify a subset of genes with altered expression patterns that were also likely targets of selection during domestication.

Results

Using a NimbleGen (Roche NimbleGen) expression array representing 32,540 genes [the filtered gene set annotation 4a.53 of the maize reference genome from Schnable et al. (25)], we collected expression profile data from 8-d-old tissue of 62 genotypes: 38 diverse maize inbred lines, 7 teosinte inbred lines, and 17 teosinte individuals sampled from three wild-collected, outcrossing populations (Table S1). Eight-day-old plants, which have cotyledons and one or two leaves (Fig. S1A), were chosen for expression profiling to minimize expression differences attributable to developmental disparity between genotypes and taxa. Of the 32,540 genes represented on the array, we first identified 19,792 genes (73,104 oligonucleotides) that showed evidence for expression based on hybridization signal above levels of random nonmaize control sequences. We eliminated an additional 26,937 probes that comparative genomic hybridization (CGH) data (26) showed to have high cross-genotype variation in hybridization signal for genomic DNA, presumably attributable to nucleotide divergence or structural variation. Our final dataset consisted of 46,167 CGHfiltered probes representing 18,242 expressed genes (1-4 probes per gene) that were Robust Multichip Average (RMA) normalized (Dataset S1) and used for subsequent analyses.

Genome-wide, the coefficient of variation of expression among lines was nearly identical in maize and teosinte (Fig. 1*A*), indicating artificial selection did not cause a transcriptome-wide

Author contributions: C.L.M., P.T., and N.M.S. designed research; R.S.-W. and R.B. performed research; R.S., M.B.H., and J.R.-I. contributed new reagents/analytic tools; R.S.-W., R.B., M.B.H., J.R.-I., C.L.M., P.T., and N.M.S. analyzed data; and C.L.M., P.T., and N.M.S. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Data deposition: The data reported in this paper have been deposited in the Gene Expression Omnibus (GEO) database, www.ncbi.nlm.nih.gov/geo (accession no. GSE30036).

¹R.S.-W. and R.B. contributed equally to this work.

²To whom correspondence should be addressed. E-mail: springer@umn.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10. 1073/pnas.1201961109/-/DCSupplemental.



Fig. 1. Variance in expression. (A) Density plots for the coefficient of variance (CoVar) for gene expression levels in all genotypes (black), maize genotypes (red), and teosinte genotypes (blue), as well as for developmental stages (green). More genes exhibit a higher CoVar across developmental stages than across diverse genotypes. (B) Relative gene expression levels for the 612 genes with significant expression differences between maize and teosinte, and used for hierarchical clustering. Genotypes were each assigned to one of five subpopulations (specified in Table S1). (C) Similar clustering is shown for the 45 genes that are differentially expressed between inbred and outcrossed teosinte. NSS (nonstiff stalk); SS (stiff stalk).

reduction in the variation of expression, nor has subsequent maize improvement resulted in vastly different expression levels among inbred lines. Expression varied more among tissues and developmental stages [data from Sekhon et al. (27)] than among different genotypes (Fig. 1*A* and Fig. S1 *B* and *C*). Among-line variance of expression differed significantly with respect to various factors, including gene conservation, genomic locations, and gene family size (Fig. S1 *D–F*).

We identified 612 genes with significantly different [posterior probability ($P_{\text{posterior}}$) of differential expression (DE) > 0.999] levels of expression in maize compared with teosinte (Table 1), with nearly half of these genes (n = 288) showing at least twofold difference in expression, with a slight bias (58.3%) toward higher expression in maize. Hierarchical clustering of the relative expression of these genes that show DE demonstrates that for some genes, there are groups of maize genotypes with expression levels similar to those in teosinte (Fig. 1*B*).

Teosinte plants normally exhibit low levels of self-fertilization (28), but modern maize breeding programs are now based on creating hybrids between inbred lines generated through controlled self-pollinations. All the maize lines included in our analyses are inbred lines. Seven of our teosinte genotypes (designated as TILs) are genetic stocks produced by multiple generations of self-fertilization. The remaining 17 teosinte samples are individuals from three wild-collected, outcrossing populations (Table S1). To investigate the effects of inbreeding on expression patterns, we compared expression between inbred and outbred teosinte. We identified 45 genes with significantly altered expression in outcrossed relative to inbred teosinte plants ($P_{\text{posterior}} > 0.999$; Table 1), with the majority of these expressed at higher levels in inbred plants (Fig. 1C). Seven of these 45 genes also showed significant differences in expression levels between maize and teosinte (Dataset S1). Among these 45 DE genes, there is significant enrichment for chitin metabolic processes and defense response genes, all of which were more highly expressed in inbred plants (Table S2). Many of these DE genes also are members of the same teosinte coexpression subnetwork that is enriched for genes annotated with biological functions in response to biotic stimulus (Table S2).

Rewiring of Transcription in Maize Relative to Teosinte. DE analyses identify individual genes with significantly different expression but may not identify changes in regulatory relationships among pairs or groups of genes. Analysis of coexpression across a set of genotypes can, however, be used to identify genes whose coregulation was altered during domestication, even if those genes' average expression relationships on a global scale, we compared the topologies of coexpression networks that were separately constructed for maize and teosinte. To quantify the coexpression within each taxon, we generated a coexpression matrix by calculating the among-genotype correlations for every pair of genes within each taxon, and these correlations are the edges of the network. We then examined the correlation between edges in the two networks;

Table 1. DE genes

Gene list	Gene no.	% 2FC	% up-regulated in maize	No. Dom or Imp candidates
Maize vs. teosinte DE	612	47	58.30	90
Teosinte inbred vs. outcrossed DE	45	95	NA	4
AEC AEC and DE	1,115 276	16 51	57.1 63.4	135 46

Dom, domestication; 2FC, 2 fold-change; Imp, improvement; NA, Not applicable.

a correlation of 1 would indicate that the patterns of coexpression in maize were identical to coexpression in teosinte. The empirical correlation between edges in the maize and teosinte networks was 0.30, which was lower than the correlation observed in all but 14 of 1,000 pairs of matrices derived from random permutations of the genotypes (Fig. 24). The rewiring of expression networks since maize-teosinte divergence is also evident from pairwise gene expression correlations, which reveal far fewer conserved gene pairs than expected based on resampled coexpression networks (Fig. 2 *B* and *C*).

To identify genes contributing the largest differences between the maize and teosinte coexpression networks, we computed an expression conservation (EC) score for each gene, a measure of the degree of similarity between a candidate gene's neighbors in the maize expression network and the same gene's neighbors in the teosinte network (*Materials and Methods*). Consistent with the correlation analyses, we observed a transcriptome-wide shift in EC score toward lower conservation, relative to the null expectation (Fig. S24). To characterize the genes showing the strongest altered expression conservation (AEC) between maize and teosinte (hereafter referred to as AEC genes), we identified a set of 1,115 genes with observed EC scores >3 SDs below the



Fig. 2. Rewiring of transcriptional networks in maize and teosinte. (*A*) Pearson correlation coefficient was determined for the full matrix correlation of maize and teosinte coexpression networks (black arrow). Only 1.4% of 1,000 pairs of networks derived from randomly permuting the genotypes exhibit lower correlations than the maize and teosinte networks. (*B*) Scatterplot shows the correlation between all gene pairs in maize (*x* axis) relative to the correlation for the same gene pair in teosinte (*y* axis). The relative density of data points in *B* was compared with the average for 1,000 bootstrap coexpression networks in C. Blue regions indicate fewer observed correlations relative to the bootstrap networks, whereas red coloration indicates an excess of actual observations, providing evidence for an enrichment of gene pairs with varying correlations in maize and teosinte.

expected EC score derived from random permutations of the genotypes.

Characterization of Genes with Altered Expression in Maize and Teosinte. The above analyses identified 612 genes with DE levels in maize and teosinte and 1,115 genes with AEC in maize and teosinte. Of these, 276 showed both DE and AEC, significantly more than expected by chance (P < 0.05). However, DE and AEC approaches identify partially distinct aspects of maize-teosinte expression changes (Fig. S2 *B* and *C*). A gene ontology (GO) analysis of genes that have reduced expression in maize relative to teosinte finds evidence for significant overrepresentation of genes related to amino acid salvage, cellular respiration, and sulfur amino acids biosynthetic processes (Fig. S34). The genes with reduced expression in maize are also overrepresented by genes that are either not located in syntenic regions in sorghum or that are maize- or grass-specific (Fig. S3 *B* and *C*). However, the majority of DE genes are syntenic with rice and sorghum.

It is possible that some of the altered expression observed in maize relative to teosinte might represent differences in development or anatomy of the two taxa. To test for this possibility, we compared the DE and AEC genes with developmental coexpression networks derived from 60 different tissues/stages of B73 (27). We did not find evidence that the DE or AEC genes were enriched in specific developmental coexpression clusters, which suggests that neither the DE nor AEC genes are the result of differences in development or morphology of maize and teosinte.

Altered Gene Expression Within Targets of Selection. Transcriptome profiling can identify genes that are either responsible for differences between species or are downstream of causative changes. To identify expression changes that are potentially responsible for differences selected during domestication, we compared DE and AEC genes with a list of genes located in genomic regions putatively selected during domestication and/or improvement (14) (Fig. 3*A*). Genes that show both DE and AEC are significantly overrepresented among genes found in these candidate regions (P < 0.05; Table 2). These genomic regions are also enriched for DE-only genes but show no significant overrepresentation of AEC-only genes (Table 2). This may provide evidence that AEC-only genes are reflecting downstream rewiring of the transcriptome but are likely not the causal sequence variants on which selection occurred during domestication.

We focused our analyses on genes in selected regions that show both DE and AEC (46 genes in Table S3) or show DE only (44 genes in Table S4), because these lists were significantly enriched for putative targets of selection. Analysis of domains present within these genes and annotation of the closest match in Arabidopsis suggest that 13 of the 90 genes may function as transcription or chromatin factors. The majority of the 90 genes show higher expression in maize (35 of 46 for the DE and AEC genes, 27 of 44 for DE-only genes) and tend to have slightly higher connectivity in the teosinte coexpression network relative to maize (Table 2). Many of the genes show substantially altered levels of connectivity in maize and teosinte (Tables S3 and S4), but this difference in connectivity does not appear to be related to the directionality of change in expression. There also is not a clear or consistent pattern in the nature of the connectivity differences shown by these genes. Coexpression subnetworks were analyzed for several of these genes to understand better how domestication affected their coregulatory relationships (Fig. 3B and Fig. S4A). These example teosinte coexpression networks include several small and moderate-sized networks. The gene used as the query for the networks (shown in red in Fig. 3B and Fig. S4A) is highly connected in teosinte. However, many of the connections for this gene are lost following domestication. Examples are also found in which parts of the teosinte coexpression network are maintained in the maize network,



Fig. 3. Analysis of genes with altered expression or conservation and targets of selection during improvement and/or domestication. (*A*) Venn diagram showing the overlap between DE genes, AEC genes, and the genes that occur in genomic regions that have evidence for selective sweeps during maize domestication or improvement (Dom/Imp genes). (*B*) Teosinte coexpression networks for three genes (GRMZM2G068436, GRMZM2G137947, and GRMZM2G375302). (*Right*) Edges that are maintained in maize coexpression networks are shown. Although the differentially expressed gene (red node) is highly connected in teosinte, most of these connections are lost in maize. However, some parts of the teosinte network are still conserved in maize. (*C*) Cross-population composite likelihood ratio test (XP-CLR) plot shows the evidence for a selective sweep that occurs on chromosome 9. The tick marks along the *x* axis represent genes, and the red tick mark indicates the gene (GRMZM2G448355) that was chosen as the candidate target of selection and is differentially expressed in maize and teosinte. The bar plot underneath the graph shows the expression levels of all maize (blue) and teosinte (red) samples. (*D*) XP-CLR plot for a large region on chromosome 5. The candidate target of selection is indicated in green and shows similar expression in maize and teosinte. Two other genes (red) exhibit DE. (*E*) Neighbor-joining tree shows the relationships among the haplotypes at GRMZM2G141858. (*Right*) Bar plot shows expression levels for each genotype; red bars indicate teosinte genotypes, and blue bars represent maize genotypes. At least one teosinte genotype (TIL15) contains the haplotype that has been selected in maize and has expression levels similar to maize genotypes.

whereas others are lost after domestication (Fig. 3*B*). It should be noted that many of these genes have unique coexpression edges in maize that are not observed in teosinte (Fig. S4*B*).

Expression data provide an opportunity to investigate further functional alterations to genes located within genomic regions that population genomic analyses identify as targets of selective

Table 2.	Genes in	selected	regions	with	evidence	for	DE (or /	AEC

Gene list	No. genes selected during dom/imp	% up-regulated in maize	Significance	% higher connected in maize	% candidates
AEC and DE (<i>n</i> = 276)	46	76	0.0002	41.3	39.1
DE only (n = 336)	44	61	0.0230	40.9	22.7
AEC only (<i>n</i> = 839)	89	54	0.1837	57.3	32.6

dom, domestication; imp, improvement.

sweeps. Many of the genomic regions identified by Hufford et al. (14) contain multiple genes, and these investigators identified the most likely target of selection within these regions by choosing the gene nearest the point in the region with the highest likelihood for selection during domestication or improvement. Many of the DE genes within selected regions (28 of 90 genes) represent the gene identified as the candidate target of selection based on population genetic metrics (Fig. 3*C*, Table 2, and Tables S3 and S4). The other 62 examples of differentially expressed genes in these regions represent examples in which the differentially expressed gene was in the selected region but was not the gene with the highest likelihood of selection (example in Fig. 3*D*). Although these genes are not located nearest the selection likelihood peak, the DE observed makes them compelling candidates nonetheless.

An examination of the expression levels of DE genes in each of the maize and teosinte genotypes reveals that although the average maize and teosinte expression levels are quite different, there is frequent overlap in the range of expression levels in maize and teosinte, such that some teosinte genotypes have expression levels similar to those found in maize or vice versa. This observation could reflect selection occurring on standing variation in teosinte, resulting in an increase in allele frequency during domestication. Conversely, the finding that some maize genotypes have expression levels more similar to teosinte could reflect examples in which the selective sweep has not been complete and some maize genotypes still retain an alternative allele. To investigate these possibilities, we compared patterns of expression of several DE genes with genetic distances calculated from the SNP data of Hufford et al. (14) (Fig. 3E and Fig. S5 A and B). The neighbor-joining trees generated using these SNP data show that the majority of maize genotypes have a very similar allele. However, there are examples in which some maize individuals group with teosinte or in which some teosinte genotypes are most similar to maize. In general, the expression level is highly correlated with the allele that is present.

Discussion

Regulatory changes have been proposed to play a major role in phenotypic evolution and some of the best-characterized morphological changes that have accompanied domestication appear to have resulted from adaptive changes at transcription factors (3, 7, 8, 24). However, the extent to which domestication, or species divergence in general, has altered the transcriptome is not well understood. By profiling the transcriptomes of 38 maize and 24 teosinte individuals, we characterized the diversity of transcriptional variation within each of these taxa, as well as the divergence of expression that has occurred during the past \sim 9,000 y since domestication. The overall coexpression networks in maize and teosinte show evidence for significant rewiring. Subsequent per gene analyses identified many examples of both DE and AEC. The analysis of coexpression networks that have substantially changed in maize relative to teosinte may assist in further understanding the molecular basis of phenotypic adaptation. In addition, the comparison of transcriptomes of wild and domesticated derivatives can begin to describe how selection on quantitative traits has affected gene expression networks.

In a search for targets of maize domestication and improvement, Hufford et al. (14) identified 484 and 695 chromosomal regions with signals of selection during domestication and improvement, respectively. A large proportion of these selected regions (58% of domestication regions and 48% of improvement regions) include multiple genes. These investigators subsequently identified the most likely target of selection within these regions by identifying the gene nearest the point in the region with the highest likelihood of selection during domestication or improvement. Analysis of maize and teosinte transcriptomes can provide a complementary approach to characterize genes within selected intervals further and to identify likely candidate targets. In some cases (28 of 90 cases), the two approaches identified the same gene, although for the remaining 62 windows of selection, the DE gene is not the gene nearest the region of the window with the greatest statistical support for being the target of selection. These windows may be examples of the limits of using population-genetic inferences alone for identifying targets of selection. Alternatively, the target of selection identified through population-genetic analyses may be correct and the altered expression may reflect a cis-regulatory variant that has "hitchhiked" along with the selected allele. Hufford et al. (14) identified 299 chromosomal regions that show evidence for selection during domestication or improvement but contain no genes from the filtered gene set. The nearest genes from the filtered gene set were compared with the 612 DE genes, and we identified five cases of DE for the genes located near selected regions (Fig. S5D). These may be examples for selection acting directly on regulatory sequences.

Although we do not have information on the functional consequences of the expression changes we detected, several DE genes are among the classic and well-studied maize genes [Schnable and Freeling (29)]. These include ae1 (amylose extender1), an1 (anther ear1), adh2 (alcohol dehydrogenase2), chn3 (chitinase3), du1 (dull endosperm1), fht1 (flavonone 3-hydroxylase 1), gln2 (glutamine syntehtase2), *lpa1* (low phytic acid1), and *zmm2* (Z. mays MADS2). Two of these genes, adh2 and zmm2, also show evidence of selection during maize domestication or improvement in the study by Hufford et al. (14). Interestingly, ael was previously identified as a target of selection during domestication [Whitt et al. (30)]. This gene is nearly twofold more highly expressed in teosinte seedlings relative to maize seedlings. The potential functional significance of additional DE genes also can be inferred based on orthology. For example, GRMZM2G448355, a domestication candidate [Hufford et al. (14)], has sequence similarity to rice OsMADS56, which is implicated in control of flowering time in rice and is located within a flowering-time QTL in maize (31).

An important limitation of the expression data we assayed is that they were collected from 8-d-old seedlings. We chose this developmental stage because all individuals, regardless of the taxon, are morphologically similar at this stage; therefore, differences we detect are likely attributable to divergence between taxa and not to comparing taxa that are at different developmental stages. An attempt to document differences in expression in visibly altered structures, such as flowers or seeds, would identify numerous expression differences that correspond to tissue differences and not necessarily expression per se. Although assaying 8-d-old seedlings allowed the isolation of comparable samples, it limited our ability to identify the targets of selection that are specifically expressed in other tissues that have been subjected to strong selective pressures. For example, only 2 (*ae1* and *su1*) of the 13 "known" targets of selection during domestication [Hufford et al. (14)] were expressed in seedling tissue; the remaining 11 genes, including *tb1* and *tga1*, are expressed specifically in other tissues (15, 16). Nevertheless, we identified differentially expressed genes in young seedling tissue that include a disproportionate amount of domestication or improvement candidates, as well as genes previously shown to play important functional roles in maize.

Despite the challenges of comparing the transcriptomes of domesticated plants and their wild ancestors, this approach can provide a detailed view of the impacts of selection on gene expression patterns. Even in young seedling tissue, before visible differences between maize and teosinte are apparent, we find evidence for significant rewiring of expression networks and find numerous differentially expressed genes. Previous studies have found that domestication targets are enriched for genes with regulatory functions (3, 21–23). This study has provided further evidence that domestication has frequently selected for regulatory variants and can also provide the basis to characterize the downstream targets of these genes.

Materials and Methods

Microarray Hybridization and Data Processing. Diverse maize inbred lines (n = 38), inbred teosinte lines (n = 7), and wild teosinte individuals (n = 17) were grown, and seedling leaf tissue was harvested 8 d after planting as previously described (26). Purified RNAs were labeled (details provided in *SI Materials and Methods*) and hybridized to a custom long-oligonucleotide microarray (GPL10846) designed by NimbleGen (Roche NimbleGen). The data were filtered to omit probes without substantial expression signal and probes with CGH variation (*SI Materials and Methods*). The raw data from the remaining 46,167 probes were renormalized using RMA (32) to provide the estimates of gene expression for 18,242 genes. Differentially expressed genes were identified using Cyber T (33) utilizing a conservative experiment-wide $P_{posterior} > 0.999$. The gene expression data generated for this study are available in the Gene Expression Omnibus database (accession no. GSE30036).

Coexpression Network Analysis. Coexpression networks were computed separately for the 38 maize expression profiles and 24 teosinte profiles by calculating Pearson correlation coefficients between all pairs of genes. Each set of correlations was then transformed using the Fisher transformation as recommended by Huttenhower et al. (34) and standard-normalized to allow for comparisons between the two networks. An EC score was calculated as the Pearson correlation coefficient between gene profiles in two coexpression networks as described by Dutilh et al. (35). The significance of differences between the maize and teosinte coexpression networks was assessed through bootstrapping, where genotypes were selected at random without replacement from the full-expression dataset, forming two groups with 38 and 24 genotypes each to match the number of genotypes in maize and teosinte subsets, respectively. Each obtained pair of subsets was used to build a coexpression network, and comparisons of the two networks were repeated on each of these bootstrapped networks. This process was repeated 1,000 times to generate null distributions for cross-network

- Gaut BS, Le Thierry d'Ennequin M, Peek AS, Sawkins MC (2000) Maize as a model for the evolution of plant nuclear genomes. Proc Natl Acad Sci USA 97:7008–7015.
- 2. Doebley J (2004) The genetics of maize evolution. Annu Rev Genet 38:37-59.
- Doebley JF, Gaut BS, Smith BD (2006) The molecular genetics of crop domestication. Cell 127:1309–1321.
- Freeling M (2009) Bias in plant gene content following different sorts of duplication: Tandem, whole-genome, segmental, or by transposition. *Annu Rev Plant Biol* 60: 433–453.
- Matsuoka Y, et al. (2002) A single domestication for maize shown by multilocus microsatellite genotyping. Proc Natl Acad Sci USA 99:6080–6084.
- Piperno DR, Ranere AJ, Holst I, Iriarte J, Dickau R (2009) Starch grain and phytolith evidence for early ninth millennium B.P. maize from the Central Balsas River Valley, Mexico. Proc Natl Acad Sci USA 106:5019–5024.
- Gross BL, Olsen KM (2010) Genetic perspectives on crop domestication. Trends Plant Sci 15:529–537.
- Purugganan MD, Fuller DQ (2009) The nature of selection during plant domestication. Nature 457:843–848.
- Vigouroux Y, 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.
- Wright SI, et al. (2005) The effects of artificial selection on the maize genome. Science 308:1310–1314.
- Yamasaki M, et al. (2005) A large-scale screen for artificial selection in maize identifies candidate agronomic loci for domestication and crop improvement. *Plant Cell* 17: 2859–2872.
- Briggs WH, McMullen MD, Gaut BS, Doebley J (2007) Linkage mapping of domestication loci in a large maize teosinte backcross resource. *Genetics* 177:1915–1928.
- Yamasaki M, Wright SI, McMullen MD (2007) Genomic screening for artificial selection during domestication and improvement in maize. Ann Bot (Lond) 100:967–973.
- Hufford MB, et al. (2012) Comparative population genomics of maize domestication and improvement. Nat Genet, 10.1038/ng.2309.
- Doebley J, Stec A, Hubbard L (1997) The evolution of apical dominance in maize. Nature 386:485–488.
- Dorweiler J, Stec A, Kermicle J, Doebley J (1993) Teosinte glume architecture 1: A genetic locus controlling a key step in maize evolution. *Science* 262:233–235.
- 17. Wang H, et al. (2005) The origin of the naked grains of maize. *Nature* 436:714–719. 18. Bomblies K, Doebley JF (2006) Pleiotropic effects of the duplicate maize FLORICAULA/
- LEAFY genes zfl1 and zfl2 on traits under selection during maize domestication. Genetics 172:519–531.
- 19. Gallavotti A, et al. (2004) The role of barren stalk1 in the architecture of maize. *Nature* 432:630–635.
- Vollbrecht E, Springer PS, Goh L, Buckler ES, 4th, Martienssen R (2005) Architecture of floral branch systems in maize and related grasses. *Nature* 436:1119–1126.

correlation, the joint edge weight distribution, and an EC score distribution for each gene. Rewired genes were selected by computing a *z* score using the gene-specific null distribution and applying a cutoff of z < -3.0. Further details of how the coexpression networks were constructed and compared are included in *SI Materials and Methods*.

Analysis of Overlap Between Expression-Derived and Sequence-Derived Selection Target Lists. Hufford et al. (14) performed sequence-based analysis to identify 3,040 genes in regions targeted during domestication and improvement, of which 1,761 genes were present in our dataset. We measured the enrichment in the sequence-derived selection targets using the hypergeometrical distribution for each of the following sets of genes: DEonly genes, AEC-only genes, and genes that are both DE and AEC.

Expression levels in DE and DE/AEC genes were compared with genetic distance between genotypes using SNP data from Hufford et al. (14). Neighbor-joining trees were constructed based on simple parsimony substitution models as implemented in the program TASSEL (version 3.0) (36).

GO Analyses. The GOslim annotation of gene lists was assessed using BiNGO (37), a Cytoscape (38) plug-in that maps overrepresented functional themes present in a given gene set onto the GO hierarchy. *P* values for enrichment of GOslim terms were calculated using a hypergeometrical distribution statistical testing method with false discovery rate correction.

ACKNOWLEDGMENTS. We thank John Doebley, who very graciously provided stocks of the inbred teosinte lines. Peter Hermanson helped with DNA isolation and microarray hybridization. The Minnesota Supercomputing Institute provided access to software and user support for data analyses. This project was supported by US Department of Agriculture Hatch funds, a seed grant from the University of Minnesota Interdisciplinary Informatics program, and Grant IOS-0922095 from the National Science Foundation (to N.M.S.). C.L.M. is also partially supported by Grant DBI-0953881 and Grant IOS-1126950 from the National Science Foundation.

- Doebley J, Lukens L (1998) Transcriptional regulators and the evolution of plant form. Plant Cell 10:1075–1082.
- Zhao Q, et al. (2008) The role of regulatory genes during maize domestication: Evidence from nucleotide polymorphism and gene expression. *Genetics* 178:2133–2143.
- Zhao Q, Weber AL, McMullen MD, Guill K, Doebley J (2010) MADS-box genes of maize: Frequent targets of selection during domestication. *Genet Res* 93(1):1–11.
- 24. Carroll SB (2008) Evo-devo and an expanding evolutionary synthesis: A genetic theory of morphological evolution. Cell 134:25–36.
- Schnable PS, et al. (2009) The B73 maize genome: Complexity, diversity, and dynamics. Science 326:1112–1115.
- Swanson-Wagner RA, et al. (2010) Pervasive gene content variation and copy number variation in maize and its undomesticated progenitor. *Genome Res* 20:1689–1699.
- Sekhon RS, et al. (2011) Genome-wide atlas of transcription during maize development. Plant J 66:553–563.
- Hufford MB, Gepts P, Ross-Ibarra J (2011) Influence of cryptic population structure on observed mating patterns in the wild progenitor of maize (Zea mays ssp. parviglumis). *Mol Ecol* 20(1):46–55.
- 29. Schnable JC, Freeling M (2011) Genes identified by visible mutant phenotypes show increased bias toward one of two subgenomes of maize. *PLoS One* 6(3):e17855.
- Whitt SR, Wilson LM, Tenaillon MI, Gaut BS, Buckler ES (2002) Genetic diversity and selection in the maize starch pathway. *Proc Natl Acad Sci USA* 99(20):12959–62.
- 31. Buckler ES, et al. (2009) The genetic architecture of maize flowering time. *Science* 325: 714–718.
- Irizarry RA, et al. (2003) Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* 4:249–264.
- Baldi P, Long AD (2001) A Bayesian framework for the analysis of microarray expression data: Regularized t-test and statistical inferences of gene changes. *Bioinformatics* 17:509–519.
- Huttenhower C, Hibbs M, Myers C, Troyanskaya OG (2006) A scalable method for integration and functional analysis of multiple microarray datasets. *Bioinformatics* 22: 2890–2897.
- Dutilh BE, Huynen MA, Snel B (2006) A global definition of expression context is conserved between orthologs, but does not correlate with sequence conservation. BMC Genomics 7:10.
- Bradbury PJ, et al. (2007) TASSEL: Software for association mapping of complex traits in diverse samples. *Bioinformatics* 23:2633–2635.
- Maere S, Heymans K, Kuiper M (2005) BiNGO: A Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. *Bioinformatics* 21: 3448–3449.
- Shannon P, et al. (2003) Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res* 13:2498–2504.

Supporting Information

Swanson-Wagner et al. 10.1073/pnas.1201961109

SI Materials and Methods

DNA Labeling and Microarray Hybridization. A custom long-oligonucleotide microarray (GPL10846) was designed by NimbleGen (Roche NimbleGen) using the 32,540 filtered maize genes predicted from the B73 reference genome (1) as described by Swanson-Wagner et al. (2). RNAs were isolated using the commercial TRIzol (catalog no. 15596026; Invitrogen) from above-ground tissue harvested 8 d after germination. RNAs were purified by lithium chloride treatment, followed by 3 M sodium acetate (0.1 vol) and 95% (vol/vol) ethanol (2.5 vol) precipitation. Purified RNAs (10 µg per sample) were reverse-transcribed and labeled according to the array manufacturer protocol (NimbleGen Arrays User's Guide: Gene Expression Analysis v3.2; Roche NimbleGen). Per sample, ~20 µg of Cy3- or Cy5-labeled RNAs was hybridized for 16–20 h at 42 °C using the NimbleGen Hybridization System (Roche NimbleGen). After hybridization, slides were washed (NimbleGen Wash Buffer Kit; Roche NimbleGen) and dried for 2 min by centrifugation. Slides were immediately scanned using the GenePix 4000B Scanner (Molecular Devices) according to the array manufacturer's protocol.

Data Processing. Array images and data were processed using NimbleScan (Roche NimbleGen) software. Briefly, images from each slide were separated into 12 subarrays and aligned to a grid to extract signal intensity for each feature on the array. Experimental integrity was verified by evaluation of the signal intensities of the sample tracking control features for each subarray. Furthermore, metrics reports were produced for each array to report the signal uniformity across the array and the intensity of known empty features, random probes, and experimental probes. A total of 78 samples provided high-quality data and were used for subsequent analyses. NimbleScan was used to generate robust multichip average (RMA) normalized (3) gene expression values from the spatially corrected probe signal intensities on a perprobe and per-gene basis. Normalized gene expression values across multiple replications (technical or biological) of the same genotype were averaged when possible. Comparisons of the distributions of signal intensity for the random sequence controls compared with the experimental gene probes on each array were used to determine a reasonable signal threshold for positive expression across all slides. There are 19,792 genes (represented by 73,104 probes) that exhibit detectable signal (significantly higher than noise measurements) in at least three genotypes. The comparative gene hybridization (CGH) ratio (genomic DNA signal relative to B73 genomic DNA) was assessed for the 73,104 probes from expressed genes in hybridizations of over 40 diverse maize and teosinte genotype (2). Probes (n = 26,937)that exhibit low CGH values in at least three genotypes were eliminated, resulting in a set of 46,167 probes that measure expression in 18,242 genes (1-4 probes per gene). The raw data from these probes were renormalized utilizing RMA to provide the estimates of gene expression used for this study. Differentially expressed genes were identified using Cyber T (4) utilizing a conservative experiment-wide posterior probability of differential expression >0.999. The gene expression data generated for this study are available in the Gene Expression Omnibus database (accession no. GSE30036).

Construction of Coexpression Networks. For the generation of coexpression networks, we arranged the expression data into a matrix *E* that had 62 genotype columns and 18,242 gene expression profile rows. Each element E_{ij} denoted the expression

level of gene *i* in genotype *j*. We split the dataset by taxon into E_M and E_T subsets of maize (18,242 × 38) and teosinte (18,242 × 24) genotypes, respectively. To build coexpression networks represented by matrices R_M and R_T , we calculated the Pearson correlation coefficient between each pair of gene expression profiles within a single subset:

$$R_{ij}^{M} = PCC\left(E_{i}^{M}, E_{j}^{M}\right)$$
$$R_{ij}^{T} = PCC\left(E_{i}^{T}, E_{j}^{T}\right)$$

for i, j = 1, ..., 18,242 and $i \neq j$. Thus, R_M and R_T were square matrices with the same dimensions, $(18,242 \times 18,242)$. Each value in those matrices represented an edge weight in the coexpression network and measured similarity between expression profiles of two genes. Even though the matrices had identical dimensions, the distribution of values in each matrix might be different because of the unequal sample size. Hence, a value from one matrix could not be compared directly with a value in the other. To enable direct comparison between R_M and R_T , we applied Fisher z transformation to both matrices as recommended by Huttenhower et al. (5). For each element r in R_M or R_T , the Fisher transformation is defined as

$$z = \frac{1}{2}\ln\frac{1+r}{1-r},$$

which guarantees that the distribution of z values is approximately normal. We normalized the distribution by subtracting the mean and dividing by the SD to obtain a standard normal distribution N(0, 1). Thus, each element in R_M and R_T represented how many SDs the corresponding pairwise similarity score was from the mean, making cross-network comparison possible. The coexpression networks were created, processed, and analyzed using the Sleipnir C++ library (6).

Comparison of Coexpression Networks. An expression conservation (EC) score was calculated as the Pearson correlation coefficient between gene profiles in two coexpression networks as described by Dutilh et al. (7):

$$EC = PCC(R_i^M, R_i^T),$$

where R_i^M and R_i^T were *i*-th rows in the matrices that represented maize and teosinte coexpression networks, respectively. For each round of bootstrapping (n = 1,000), genotypes were selected at random without replacement from the full expression dataset forming two groups with 38 and 24 genotypes each to match the number of genotypes in maize and teosinte subsets, respectively. Each obtained pair of subsets was then used to build a coexpression network. Because each group had a random mix of maize and teosinte accessions, any variations between coexpression networks would exist mainly as a result of random variation and the difference in group size. The same analysis was performed on each corresponding pair of the bootstrapped coexpression networks, including the calculation of matrix correlation, edge weight distribution, and EC score distribution. Null expectations for EC score and joint edge weight distribution were calculated as averages across all bootstraps. We selected rewired genes using a z score calculated for each EC value as follows:

$$z = \frac{EC - \mu}{\sigma},$$

where μ and σ were the mean and SD of the gene's EC scores obtained from the bootstrapping analysis. The *z* score of rewired genes was required to be below -3.0.

Gene Annotations and Gene Ontology Analyses. The maize-specific genes and gene families were identified based on homolog clustering with annotated genes of rice, sorghum, and *Arabidopsis* using the method of Vilella et al. (8) as previously described (2). Pa-

- 1. Schnable PS, et al. (2009) The B73 maize genome: Complexity, diversity, and dynamics. Science 326:1112–1115.
- Swanson-Wagner RA, et al. (2010) Pervasive gene content variation and copy number variation in maize and its undomesticated progenitor. *Genome Res* 20:1689– 1699.
- 3. Irizarry RA, et al. (2003) Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* 4:249–264.
- Baİdi P, Long AD (2001) A Bayesian framework for the analysis of microarray expression data: Regularized t-test and statistical inferences of gene changes. *Bioinformatics* 17: 509–519.
- Huttenhower C, Hibbs M, Myers C, Troyanskaya OG (2006) A scalable method for integration and functional analysis of multiple microarray datasets. *Bioinformatics* 22: 2890–2897.

ralogous clusters were defined as two or more genes belonging to the same gene family that were separated on a chromosome by no more than two nonparalogous intervening genes. Syntenic mapping of maize genes to rice and sorghum was previously described (2). The gene ontology (GO) GOslim annotation of genes that were affected by structural variation was assessed using BiNGO (9), a Cytoscape (10) plug-in that maps overrepresented functional themes present in a given gene set onto the GO hierarchy. *P* values for enrichment of GOslim terms were calculated using a hypergeometrical distribution statistical testing method with false discovery rate correction.

- Huttenhower C, Schroeder M, Chikina MD, Troyanskaya OG (2008) The Sleipnir library for computational functional genomics. *Bioinformatics* 24(13):1559–1561.
- Dutilh BE, Huynen MA, Snel B (2006) A global definition of expression context is conserved between orthologs, but does not correlate with sequence conservation. BMC Genomics 7:10.
- Vilella AJ, et al. (2009) EnsemblCompara GeneTrees: Complete, duplication-aware phylogenetic trees in vertebrates. *Genome Res* 19(2):327–35.
- Maere S, Heymans K, Kuiper M (2005) BiNGO: A Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. *Bioinformatics* 21: 3448–3449.
- Shannon P, et al. (2003) Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res* 13:2498–2504.



Fig. 51. (*A*) Images of sampled plants. Several images of 8-d-old plants immediately before harvesting for RNA isolation. (*B*) Coefficient of variance (CoVar) for gene expression levels was compared for 13,789 genes in a survey of different genotypes (*y* axis) and different developmental stages (*x* axis). (*C*) CoVar for gene expression levels in all maize genotypes (*y* axis) and all teosinte genotypes (*y* axis) was contrasted. The red data points indicate the 612 genes that are differentially expressed in maize relative to teosinte. The star-shaped data points indicate genes that were identified as domestication and/or improvement candidates by Hufford et al. (1). (*D*) Box plot is shown for the coefficient of variance for gene expression levels in all maize and teosinte genotypes analyzed for different subsets of genes. The genes are divided according to the depth of their homology. Genes with no homologs in maize or other species are classified as "NULL." Genes that have multiple family members in maize but are not detected in other species are classified as "*Zea mays*." Genes that have a sorghum homolog but no homologs in other grasses are classified as "Andropogoneae." Genes that have homologs in multiple grass species are classified as "Poaceae," and genes with homologs in other plant genomes are classified as "Magnoliophyta." (*E*) Box plot of gene expression variance is shown for genes that are located in syntenic (syn) and nonsyntenic (non_syn) genomic positions relative to other grass genomes. (*F*) Box plot shows gene expression variance for genes according to the gene family size in the maize genome.

1. Hufford MB, et al. (2012) Comparative population genomics of maize domestication and improvement. Nat Genet, 10.1038/ng.2309.



Fig. 52. (A) Distribution of EC scores among maize and teosinte. The distribution of EC scores (blue histogram) is shown relative to the distribution of EC scores observed in bootstrap samples (red line). The enrichment for reduced EC scores relative to the bootstrap control suggests that many genes show substantial variation for their coexpression with other genes. (*B* and C) Interaction among genes discovered by expression level or conservation contrasts in maize and teosinte. (*B*) Relative expression levels in maize and teosinte are plotted for all genes. The color of the symbols indicate whether they were identified as differential expression (DE; red), altered expression conservation (AEC; blue), both AEC and DE (black), or not found in either list (gray). Note there are numerous genes with altered EC scores (blue) that exhibit no evidence for altered expression levels because they plot near the center of the distribution. (*C*) EC (*y* axis) scores are plotted for each of 18,242 genes. (*Upper Right*) It is noteworthy that there are many DE genes that do not exhibit evidence for low EC scores (data points plotted).



Fig. S3. Characterization of genes with altered expression levels or expression conservation in maize and teosinte. (A) BiNGO was used to assess overrepresentation of GO terms in genes that are more highly expressed in teosinte than in maize. (B) Homology of each gene to other species was determined. Genes were assigned as Zea_single if they are single copy in the maize genome and do not have homologs in other species, Zea_multi if they are multicopy but only detected in maize, Andropogoneae if they have homologs in sorghum but not in rice, Poaceae if they have homologs in rice and sorghum, or Magnoliophyta if they have homologs in other plants like *Arabidopsis* or poplar. The differential expression (DE) and DE/altered expression conservation (AEC) categories have more Zea_multi copy genes than expected. A comparative analysis of the genomic location for each gene in maize and sorghum was used to assign each gene as syntenic or nonsyntenic. The DE and DE/AEC genes are significantly (*P* < 0.05) enriched for nonsyntenic positions.



Fig. 54. (*A*) Partial conservation of coexpression networks following domestication is shown. Teosinte coexpression networks are shown for two highly connected genes. (*Right*) Edges that are maintained within maize coexpression networks are shown. Although the differentially expressed gene (red node) is highly connected in teosinte, most of these connections are lost in maize. However, some parts of the teosinte network are still conserved in maize. (*B*) Coexpression networks in maize and conserved edges in teosinte are shown. For the genes that were used in *A* and Fig. 3, we generated the maize coexpression network and then show the edges that are conserved in teosinte. This demonstrates that there are edges present in maize following domestication that were not apparent in teosinte.



Fig. S5. Comparison of sequence haplotype and expression levels. (*A*–*C*) Neighbor-joining trees are created for three genes that are located within genomic regions that experienced selective sweeps. These represent examples of maize genotypes that contain an allele similar to teosinte genotypes, or vice versa. (*Right*) Bar plot shows the relative expression levels for each genotype. The blue bars indicate maize genotypes, and the red bars indicate teosinte. The scale bar indicates expected substitutions per site. (*D*) Example of a differentially expressed gene near a genomic region with evidence of selection. XP-CLR (Cross-population composite likelihood ratio test). Some of the regions identified by Hufford et al. (1) as putative targets of selection did not contain any annotated genes. We found examples in which genes near these regions are differentially expressed in maize and teosinte. In this example, the gene shows lower expression in teosinte and selection may have acted on regulatory regions that affect this gene.

1. Hufford MB, et al. (2012) Comparative population genomics of maize domestication and improvement. Nat Genet, 10.1038/ng.2309.

Table S1. List of genotypes and classifications

PNAS PNAS

			Group				Structure-based	
Genotype	Maize or teosinte	Cross-type	assignment	Tissue source	NAM?	Seed accession no.	assignments	Hufford et al. (1)
A680	Maize	Inbred	SS	Individual		Ames 23503	SS	
B73	Maize	Inbred	55	Pool	NAM	PI 550473	SS	Yes
B79	Maize	Inbred	MIX	Individual		PI 608766	NSS	i es
B84	Maize	Inbred	55	Individual		PI 608767	SS	
B07	Maize	Inbred	NSS	Pool	ΝΔΜ	PI 564682	NSS	Vec
CMI 102	Maizo	Inbred	Trop	Pool	NAM	Amor 27091	Trop	Vor
	Maize	Inbred	Trop	Pool		Ames 27081	Trop	Vec
CIVILZZO	Maize	Inbred	Ттор	Pool			пор	Yes
CIVIL247	Maize	Inbred	Ттор	Pool			Tron	Yes
CIVIL277	Maize	Inbred	Тгор	POOL		PI 595550	Ттор	Yes
CIVIL322	Maize	Inbred	тор	POOL	NAW	Ames 27096		res
CIVIL333	Maize	Inbred	Trop	POOL	NAM	Ames 27101	Trop	Yes
CML52	Maize	Inbred	Irop	Pool	NAM	PI 595561	Irop	Yes
Hp301	Maize	Inbred	Рор	Pool	NAM	PI 58/131	Irop	Yes
1205	Maize	Inbred	MIX	Individual		NSL 658/1	NSS	
II14H	Maize	Inbred	Sweet	Pool	NAM	Ames 27118	Other	Yes
Ki11	Maize	Inbred	Trop	Pool	NAM	Ames 27124	Trop	Yes
Ki3	Maize	Inbred	Trop	Pool	NAM	Ames 27123	Trop	Yes
Ky21	Maize	Inbred	NSS	Pol	NAM	Ames 27130	NSS	Yes
LH1	Maize	Inbred	PVP	Individual		PI 644101	Other	
LH82	Maize	Inbred	PVP	Individual		PI 601170	NSS	
M162W	Maize	Inbred	Trop	Pool	NAM	Ames 27134	Trop	Yes
M37W	Maize	Inbred	MIX	Pool	NAM	Ames 27133	Trop	Yes
Mo17	Maize	Inbred	NSS	Individual		PI 558532	NSS	Yes
Mo18W	Maize	Inbred	NSS	Pool	NAM	PI 550441	Trop	Yes
NC358	Maize	Inbred	Trop	Pool	NAM	Ames 27175	Trop	Yes
Oh43	Maize	Inbred	NSS	Pool	NAM	Ames 19288	Other	Yes
Oh7B	Maize	Inbred	NSS	Pool	NAM	Ames 19323	NSS	Yes
P39	Maize	Inbred	Sweet	Pool	NAM	Ames 28186	Other	Yes
PHG35	Maize	Inbred	PVP	Individual		PI 601008	NSS	
PHG39	Maize	Inbred	PVP	Individual		PI 600981	NSS	
PHG47	Maize	Inbred	PVP	Individual		PI 601320	NSS	
PHG84	Maize	Inbred	PVP	Individual		PI 601321	NSS	
PHJ40	Maize	Inbred	PVP	Individual		PI 601322	NSS	
PHZ51	Maize	Inbred	PVP	Individual		PI 601322	NSS	
Tx303	Maize	Inbred		Pool	ΝΔΜ	Ames 19327	Trop	Yes
Tzi8	Maize	Inbred	Trop	Pool	NAM	PI 506246	Trop	Yes
W22	Maize	Inbred	NSS	Individual		NSI 30053	NSS	Yes
W/f9	Maize	Inbred	NSS	Individual		Ames 19793	NSS	i es
14 29 01	Teosinte	Outcrossed	TO	Individual		Ames 21809	TO	
142511	Teosinte	Outcrossed	то	Individual		Ames 21809	то	
142912	Teosinte	Outcrossed	то	Individual		Ames 21809	то	
	Teosinte	Outcrossed	то	Individual		Ames 21805	то	
142005	Teosinte	Outcrossed	то	Individual		Ames 21805	то	
142953	Teosinte	Outcrossed	то	Individual		Ames 21809	то	
IA29P0	Teosinte	Outcrossed	TO	Individual		Ames 21009	TO	
	Teosinte	Outcrossed	TO	Individual		Ames 21810	TO	
IA31P2	Teosinte	Outcrossed	TO	Individual		Ames 21810	TO	
IA3 IP3	Teosinte	Outcrossed	10	Individual		Ames 21810	10	
IA31P4		Outcrossed	10	Individual		Ames 21810	10	
IA31P5	Teosinte	Outcrossed	10	Individual		Ames 21810	10	
IA31P6	leosinte	Outcrossed	10	Individual		Ames 21810	10	
IA36P1	leosinte	Outcrossed	10	Individual		Ames 21814	10	
IA36P2	Teosinte	Outcrossed	то	Individual		Ames 21814	то	
IA36P3	Teosinte	Outcrossed	то	Individual		Ames 21814	то	
IA36P5	Teosinte	Outcrossed	то	Individual		Ames 21814	то	
IA36P6	Teosinte	Outcrossed	то	Individual		Ames 21814	то	
TIL1	Teosinte	Inbred	TI	Individual		John Doebley*	TI	Yes
TIL11	Teosinte	Inbred	TI	Individual		John Doebley*	TI	Yes
TIL14	Teosinte	Inbred	TI	Individual		John Doebley*	TI	Yes
TIL15	Teosinte	Inbred	TI	Individual		John Doebley*	TI	Yes
TIL17	Teosinte	Inbred	ТІ	Individual		John Doebley*	ТІ	Yes
TIL6	Teosinte	Inbred	ТІ	Individual		John Doebley*	ТІ	Yes
TIL9	Teosinte	Inbred	ТІ	Individual		John Doebley*	ТІ	Yes

MI, mixed genetic background; NAM, nested association mapping parent; NSS, nonstiff stalk; Pop, popcorn; PVP, plant varietal protection; SS, stiff stalk; TI, teosinte inbred; TO, teosinte out-crossed; Trop, tropical.

*Gift from John Doebley (University of Wisconsin, Madison WI).

1. Hufford MB, et al. (2012) Comparative population genomics of maize domestication and improvement. Nat Genet, 10.1038/ng.2309.

P_inb/		0.18	0.29	0.37		0.38	2.2	0.38	0.43		07.0	9				0.49	0.50	0 50		3.14		3.72			4.58		4.72		4.82		10.4	CP.4		5.03	5.13		5.70	
Short_description		Glycine-rich protein		Eukaryotic aspartyl	protease family	protein Givrosvi hvidrolase	superfamily protein	-	Pyruvate	orthophosphate	dikinase Plaan santaining	r -roup contanting purdeoside	triphosphate	hydrolases	superfamily protein		Cytochrome	p450 78a9 Hich-mobility	aroup B4	GST family protein					FAD-binding berberine family protein		Leucine-rich repeat		Peroxidase	superfamily protein					Chitinase A		Dynein light chain	tvpe 1 familv protein
ArabidopsisID		AT5G46730.1		AT2G03200.1		AT3G06510 2			AT4G15530.5		AT1/060640.1						AT3G61880.2	AT7617560 1		AT3G62760.1	L				AT4G20820.1		AT1G73066.1	L	AT4G33420.1	L		1.00202021A			AI5024090.1	L	AT3G16120.1	
mclteosinte_GO		Signal transduction																		Molecular_function,	response to biotic stimulus, extracellula	region Molecular_function,	response to biotic stimulus. extracellula	region	Molecular_function, response to biotic stimulus. extracellula	region	Molecular_function,	stimulus, extracellula	region Mmolecular function,	response to biotic stimulus, extracellula	region	intolecular_runction, response to biotic stimulus, extracellula	region		Molecular_tunction, response to biotic	stimulus, extracellula	region Molecular_function,	reconstruction to biotic
Mclt z score		2.9871	1.7396								רטר ר	7707.7				1.2116		1 1298		3.1667		3.1667			3.1667		3.1667		3.1667			3.100/		5.2506	3.166/		3.1667	
mclt pval	Ī	0	0.055								c	5				0		c	•	0		0			0		0		0		c	þ		0	0		0	
o. Mclteo z score		0.0665111	0.7962056								CC11000.0					-0.003823		0 1216169		0.289182		0.289182			0.289182		0.289182		0.289182			7916270		0.0758775	0.28182		0.289182	
mcl teo gen n		4	4	2		ç	I				~	t				108		465	1	110		110			110		110		110					'n	011		110	
degree	b	œ	56	22		15	P	1	70		6	2				53	27	121	į	76		122			119		58		106		ļ	40		42	5		62	
Mdegree T	'n	6	110	16		1	1	0	-		ç	2				23	0	ų	•	-		ß			ъ		m		4		Ċ	N		-	F		9	
EC_zScore		-0.245	-0.762	-5.824		0.037	1000	0.377	-0.483		1 1 1 1	±				0.068	-0.637	-3 266	202	-0.605		-0.921			-1.49		-2.237		-0.442			-0.597		0.611	-3.659		-2.353	
Ű		0.125	0.55	-0.335		785.0	000	0.283	0.264			117.0				0.268	0.091	0 346	2	0.307		0.34			0.26		0.008		0.064			601.0		0.342	0.169		0.394	
Teo uster		776	006	3,022 -		550		s	S		00					1	S	ſ	n	10		10			10		10		10			2		422	01		10	
laize uster cl		103	S	178		166 3	8	s	866		E 0.7					7	S	784		77		506			154		350		557		f	7/0		,911	136		994	
'i avq DE EC c	0	319 No No	540 No No	971 No Yes				1,495 No No	1,269 No No							4,768 No No	1,985 No No	3 674 Ves No 2		1,238 No No		230 No No			405 No No		335 No No		722 No No			230 NO NO 235		1,050 No No 2	/,/1/ Yes Yes		733 Yes No	
avg P	0	,747	,882	,656		700		,889	,925		c01					1,717	;,965	500	201	394		62			88		71		150		[6		209	۲ <u>0</u> ۲,		129	
avg Pc	0	826 1	811 1	778 2		185 5	3	542 3	843 2		с С	, <u>,</u>				413 9	224 3	7 474 7		774		180			229		186		303		Ĺ	20		604	1 450		547	
Chr M		-	-	4 ,		с С	1 1	3 2,	8 1,		, ,	i t				8 7,	2, 2,	с С	ĩ	10		2			ი		ъ		10			-		ы	ο Γ		m	
Gene ID		GRMZM2G118173	GRMZM2G469224	GRMZM2G444623		AC148152 3 EG008		GRMZM2G119638	GRMZM2G097457		3476152647442					GRMZM2G154508	GRMZM2G034471	GRM7M7G162284		GRMZM2G132093		GRMZM2G056329			GRMZM2G080839		GRMZM2G174128		AC205413.4 FG001					GRMZM2G073114	GKMZM26423802		GRMZM2G016435	

PNAS PNAS

Table S2. Genes that are differentially expressed in inbred and outcrossed teosinte

Table S2. Coi	٦t.																		
Gene ID	Chr	M_avg P	o_avg }	Pi_avg D	Ĕ	Maize Cluster	Teo cluster	EC	EC_zScore	Mdegree [.]	Tdegree m	icl_teo_gen no.	Mclteo z score m	clt_pval	Mclt z score mclteosint	GO Arabidop:	sisID Short_	description P.	inb/
GRMZM2G415529	2	465	191	1,127 N	o No	48	10	0.265	-0.486	20	11	110	0.289182	0	3.1667 Molecular_fun response to stimulus, ext region	tion, AT1G669! biotic acellular	60.1 Pleiotropic resistanc	: drug ce 11	5.88
GRMZM2G143153	m	773	309	1,826 N	0 N O	0 154	10	0.129	-1.626	٢	78	110	0.289182	0	3.1667 Molecular_fun response to l stimulus, ext	tion, viotic acellular			5.91
GRMZM2G135862	Ŋ	356	100	610 N	o No	350	10	0.049	-2.706	-	99	110	0.289182	0	3.1667 Molecular_fun response to l stimulus, ext region	tion, AT4G3618 biotic acellular	80.1 Leucine-ric receptor protein familv p	:h -like kinase rotein	6.07
GRMZM2G023847	-	801	353	2,274 N	o No	0 1,580	10	0.137	-0.679	ы	67	110	0.289182	0	3.1667 Molecular_fun response to stimulus, ext region	tion, AT3G6028 biotic acellular	30.1 Uclacyanin	m	6.45
GRMZM2G181227	10	2,716	771	4,999 N	o No	9 459	10	0.296	-0.525	22	52	110	0.289182	0	3.1667 Molecular_fun response to stimulus, ext region	tion, AT5G6594 biotic acellular	40.1 β-Hydroxyi hydrolas	isobutyryl-CoA ie 1	6.48
GRMZM2G039009	Q	163	96	660 N	o No	48	10	0.36	0.65	21	115	110	0.289182	0	3.1667 Molecular_fun response to stimulus, ext region	tion, AT1G196/ biotic acellular	40.1 Jasmonic a methyltr	acid carboxyl ansferase	6.87
GRMZM2G089506	თ	166	62	434 N	0 N	0 136	10	0.284	0.081	σ	112	110	0.289182	0	3.1667 Molecular_fun response to stimulus, ext region	tion, AT5G107 biotic acellular	70.1 Eukaryotic protease protein	aspartyl e family	7.03
GRMZM2G176798	4	444	294	2,361 N	0 No	0 154	10	0.262	0.507	4	65	110	0.289182	0	3.1667 Molecular_fun response to stimulus, ext region	tion, AT5G391' biotic acellular	10.1 RmlC-like o superfan protein	nily	8.02
GRMZM2G131099	m	381	70	567 N	o No	0 136	10	0.077	-2.438	23	114	110	0.289182	0	3.1667 Molecular_fun response to l stimulus, ext region	tion, biotic acellular			8.13
GRMZM2G060630 GRMZM2G174994	10	135 345	232 106	1,913 Y. 887 N	es Ye o No	- 48 48	5 10	-0.21 0.363	-6.06 0.113	190 25	14 93	110	0.289182	0	3.1667 Molecular_fun response to l stimulus, ext	ttion, biotic acellular			8.25 8.37
GRMZM2G016922	-	233	59	518 N	0 N	9 48	10	0.39	-1.148	18	96	110	0.289182	0	3.1667 Molecular_fun response to l stimulus, ext region	tion, AT1G7940 Diotic acellular	50.1 Terpenoid prenyltra superfan	cyclases/protein ansferases nily protein	8.77
GRMZM2G057093	-	267	87	785 N	o No	2,180	10	0.189	-0.561	2	69	110	0.289182	0	3.1667 Molecular_fun response to stimulus, ext	tion, biotic acellular			9.02
GRMZM2G399530	10	330	110	1,090 N	0 N	0 135	10	0.213	-0.042	10	103	110	0.289182	0	3.1667 Molecular_fun response to l stimulus, ext region	tion, AT5G2496 biotic acellular	50.1 Cytochrom family 7 polypep	ne P450, 1, subfamily A, tide 14	9.89 9.80

Table S2. Co	Ľ.																			
Gene ID	Chr	M_avg	Po_avg	Pi_avg	DE	EC clu	aize Te ster clus	eo ster	EC	C_zScore N	1degree	Ldegree m	d_teo_gen no.	Mclteo z score	mclt_pv8	Mclt z I score	mclteosinte_GO	ArabidopsisID	Short_description	P_inb/ P_out
GRMZM2G117971	4	8,002	1,339	13,299	Yes `	Yes	136	10	0.168	-3.008	7	58	110	0.289182	0	3.1667 N	Aolecular_function, response to biotic stimulus, extracellular	AT3G04720.1	Pathogenesis-related 4	9.93
GRMZM2G337594	σ	761	74	783	Yes I	2	136	10	0.044	-2.684	10	68	110	0.289182	0	3.1667 N	region Aolecular_function, response to biotic stimulus, extracellular			10.61
GRMZM2G065585	m	817	168	1,829	No	° N	136	10	0.278	-0.705	10	102	110	0.289182	0	3.1667 N	region Aolecular_function, response to biotic stimulus, extracellular	AT5G56590.1	O-glycosyl hydrolases family 17 protein	10.86
GRMZM2G170857	4	432	183	2,060	No	°N N	154	10 (0.23	0.482	4	67	110	0.289182	0	3.1667 N	function, folecular_function, response to biotic stimulus, extracellular	AT5G39110.1	RmlC-like cupins superfamily protein	11.24
GRMZM2G161521	4	346	98	1,172	No	N N	154	10	0.306	0.067	Ŀ	155	110	0.289182	0	3.1667 N	Allecular_function, response to biotic stimulus, extracellular			11.97
GRMZM2G015933	7	345	58	758	No	°N N	135	10 (0.287	-0.374	17	107	110	0.289182	0	3.1667 N	Aolecular_function, response to biotic stimulus, extracellular	AT1G35710.1	Protein kinase family protein with leucine-rich repeat domain	13.18
GRMZM2G127087	10	220	62	1,129	- N	N	48	10	0.19	0.313	4	109	110	0.289182	0	3.1667 N	Adlecular_function, response to biotic stimulus, extracellular region	AT1G70080.1	Terpenoid cyclases/protein prenyltransferases superfamilv protein	18.10
GRMZM2G179896	10	432	68	1,264	- N	N	48	10	0.295	0.282	24	86	110	0.289182	0	3.1667 N	Aolecular_function, response to biotic stimulus, extracellular region			18.72
GRMZM2G106177	10	811	63	1,225	Yes I	No 2,4	414	10	0.295	-2.512	31	80	110	0.289182	0	3.1667 N	Aolecular_function, response to biotic stimulus, extracellular	AT1G10030.1	Homolog of yeast ergosterol28	19.42
GRMZM2G028306	10	432	133	2,705	- 2	N	48	10	0.302	0.247	12	111	110	0.289182	0	3.1667 N	region folecular_function, response to biotic stimulus, extracellular	AT1G70080.1	Terpenoid cyclases/protein prenyltransferases superfamily protein	20.38
GRMZM2G154870	-	130	49	1,024	No	No 1,!	581	10	0.199	0.00	σ	95	110	0.289182	0	3.1667 N	region Aolecular_function, response to biotic stimulus, extracellular region	AT5G57220.1	Cytochrome P450, family 81, subfamily F, polypeptide 2	20.75
GRMZM2G006853	-	1,790	89	4,277	- No	о N	48	10	0.308	-1.259	21	68	110	0.289182	0	3.1667 N	Aolecular_function, response to biotic stimulus, extracellular region	AT4G11650.1	Osmotin 34	47.80

	Table S3.	Differential expression	genes with altered e	expression conservation	that are located in re	gions with selective swee	ps
--	-----------	-------------------------	----------------------	-------------------------	------------------------	---------------------------	----

Gene ID	Chr	Maize avg	Teosinte avg	Maize degree	Teosinte degree	Dom/Imp status	Interpro annotation	Best <i>Arabidopsis</i> match
AC204921.4_FG011	6	5,315	3,868	2	12	Dom candidate	Histidine biosynthesis	
AC211164.5_FG004	2	752	352	43	32	Dcand	Haem peroxidase	AT5G67400.1
AC233910.1_FG006	2	2,980	4,356	257	7	Iregion		
AC234169.1_FG003	5	2,358	1,658	77	7	Dregion	DNA glycosylase	
AC235544.1_FG005	9	2,938	4,110	3	13	Iregion		
GRMZM2G000353	6	254	85	12	2	Imp candidate	Zinc finger, RING-CH-type	AT1G20823.1
GRMZM2G000593	3	1,747	315	6	9	Dom candidate	MATE (multidrug efflux protein)	
GRMZM2G003165	7	467	239	9	2	Dom region	Fasciclin domain	AT4G12730.1
GRMZM2G014091	8	10,627	7,197	13	33	Dom region	Uncharacterized	AT2G18740.1
GRMZM2G014720	3	3,951	2,840	8	10	DI_region	BCL-2-associated athanogene 4	AT3G51780.1
GRMZM2G015892	1	5,371	1,561	20	10	Imp candidate	Tryptophan synthase	AT4G02610.1
GRMZM2G016657	8	139	470	30	8	Dom candidate	Rhicadhesin receptor (cupin domain)	AT1G09560.1
GRMZM2G018849	9	335	94	0	2	Imp candidate	Uncharacterized	AT3G27230.1
GRMZM2G021498	1	3,231	2,271	5	7	Dom region	Zinc finger, RING-CH-type	AT3G63530.1
GRMZM2G022538	4	14,211	6,511	2	5	Dom candidate	Uncharacterized	
GRMZM2G028766	4	3.833	2,488	34	34	Dom region	Zinc finger, C2H2-type	AT1G04850.1
GRMZM2G042789	8	340	142	0	15	Imp region	Proteinase inhibitor	
GRMZM2G044180	9	229	56	1	8	Imp region	Protein kinase-like	AT1G73660.1
GRMZM2G048928	2	3216	1.852	289	2	Dom region	Tropomyosin	AT1G24560.1
GRMZM2G049346	1	1.538	1.057	2	5	Dcand	Zinc finger, RING-CH-type	AT3G19950.1
GRMZM2G049510	3	799	91	4	14	Dom candidate	KIP1-like	AT1G09720.1
GRMZM2G066555	5	5.339	3.692	20	6	Dom/Imp region	Brix domain	AT1G63780.1
GRM7M2G068323	6	1,296	624	3	1	Imp region	Pentatricopeptide repeat	AT1G63330.1
GRM7M2G068436	6	1,998	941	14	18	Imp region	Zinc finger, RING-CH-type	AT1G02610.1
GRM7M2G075315	10	1,504	624	73	35	Iregion	Line miger, mile en type	
GRM7M2G077673	7	5,669	3 491	11	7	Dregion	β-Ureidopropionase	AT5G64370.1
GRM7M2G102183	2	2 043	1 010	1	84	Dregion	Malate synthase	AT5G03860 1
GRM7M2G104999	2	8 640	4 572	5	78	Dregion	Uridine-ribohydrolase 1	AT2G36310 1
GRM7M2G119248	1	1 130	702	6	35	Imp candidate	Bromodomain transcription factor	AT2G03667 1
GRM7M2G119483	8	1,130	748	5	4	Imp candidate	Heat shock protein Dnal	AT4G28480 1
GRM7M2G141152	10	495	1 437	3	0	Dom region	Uncharacterized	711102010011
GRM7M2G141858	5	1 831	1,457	9	15	Imp region	Plant linid transfer protein	
GRM7M2G172399	10	1 153	733	2	8	Imp region	RNA recognition motif	
GRM7M2G177349	2	239	2 591	0	7	Dcand	Hydroxycinnamoyl-CoA	AT5G48930 1
GIUIZUZGITTS45	2	235	2,551	Ū	,	Deana	shikimate/quinate	A13040330.1
							hydroxycinnamoyl transferase	
GRM7M2G314707	2	1 750	868	4	з	Iregion	TRAE-like superfamily protein	AT3G11950 1
GRM7M2G315902	2	448	1 571	1	5	Imp region	Uncharacterized	AT4G35520.1
GRM7M2G320591	7	73	333	17	9	Dom region	Uncharacterized	7114355520.1
GRM7M2G32328	6	2 9 2 1	5 081	2	35	Dom region		AT/G28220 1
GRM7M2G324886	2	12 252	8 900	1/10	9	Icand	SGS domain-containing protein	AT1G30070 2
GRM7M2G359952 (zmm2)	8	433	1 016	145	3	Imn candidate	MADS-box	AT4G18960 1
GRM7M2G371795	5	2 7 1 8	1,010	2	13	Imp candidate	Orphan nuclear recentor NOR1 type	A14010500.1
GRM7M2G20501755	ר ר	6 1/0	د رد, ر د ۱, ۵	∠ و1	2/	Dom/Imp region	Uncharacterized	
GRN7N2G1/9255	∠ ۵	0,149	4,205 176	וס כ	24 17	Deand	"Transcription factor K box"	
GRN7N2C4440333	2	1 016	1 20	2	1/	Dom rogion	NAD(P) hinding	
	ש כ	1,010	1,001	כ סר	15	Dom region	Uncharacterized	
	2 1	212	074	20	2 12	Dom region	PDM1 interacting protein 4	AT2C2E070 1
	1	3,082	1,518	9	13	Dom region	Krivit interacting protein 4	A13G25070.1

avg, average; Chr, chromosome; Dom, domestication; Imp, improvement.

Table S4. Differential expression (but not altered expression conservation) gene levels that are located in regions with selective sweeps

Gene ID	Chr	Maize avg	Teosinte avg	Maize degree	Teosinte degree	Dom/Imp status	Interpro annotation	Best <i>Arabidopsis</i> match
AC193754.3_FG008	2	356	178	64	19	Dregion	MUTS homolog 2	AT3G18524.1
AC204530.4_FG005	1	10,510	6,913	6	69	Dregion	Haloacid dehalogenase-like hydrolase	
AC234521.1_FG003	2	1,021	1,950	83	76	Icand		
GRMZM2G005732	7	1,467	2,612	3	4	lcand		AT5G02810.1
GRMZM2G022310	7	766	464	0	156	DI_region	UDP-glycosyltransferase superfamily protein	AT3G45100.1
GRMZM2G025340	7	3,019	1,636	2	7	Dregion	P-loop containing nucleoside triphosphate hydrolase	AT5G61460.1
GRMZM2G025648	1	6,946	4,586	51	3	Iregion		AT1G73885.1
GRMZM2G026793	2	403	768	1	1	Iregion		AT2G43710.2
GRMZM2G027673	4	990	670	3	14	Dregion	Plant stearoyl-acyl-carrier-protein desaturase family protein	AT2G43710.2
GRMZM2G033356	4	2,086	1,451	20	1	Dcand	Basic helix-loop-helix DNA-binding superfamily protein	AT1G12860.1
GRMZM2G039016	5	457	1,153	12	58	Dregion	Pentatricopeptide repeat	AT1G74580.1
GRMZM2G071959	7	5,637	3,964	387	10	Dregion	Histone superfamily protein	AT3G45980.1
GRMZM2G075562	5	2,364	1,525	177	10	Dregion	Zinc finger, B-box, CCT domain (CONSTANS-like 9)	AT3G07650.1
GRMZM2G077632	2	7,538	3,310	34	2	Dcand	GTP1/OBG family protein	AT5G18570.1
GRMZM2G082037	6	384	218	26	4	Iregion	UDP-glucosyl transferase 85A4	AT1G78270.1
GRMZM2G083755	1	364	216	6	4	Dregion	Frataxin homolog	AT4G03240.1
GRMZM2G083984	2	628	1,087	16	100	Dregion	Mitochondrial processing peptidase alpha subunit	AT3G16480.1
GRMZM2G091456	1	993	567	7	3	Icand	FAD/NAD(P)-binding oxidoreductase family protein	AT1G58440.1
GRMZM2G098346	4	13,086	8,182	5	12	Dcand	Alcohol dehydrogenase 1	AT1G77120.1
GRMZM2G099891	5	3,872	2,543	26	125	Dregion	Chloroplast outer envelope protein 37	AT2G43950.1
GRMZM2G105229	10	567	316	11	16	Dregion	Pentatricopeptide repeat	AT4G02820.1
GRMZM2G107211	3	421	767	2	18	Dregion	Hydroxycinnamoyl-CoA shikimate/quinate hydroxycinnamoyl transferase	AT5G48930.1
GRMZM2G108537	6	1,545	841	1	10	Dregion	Nodulin MtN21 /EamA-like transporter family protein	AT1G21890.1
GRMZM2G128665	1	4,435	3,028	30	27	Dcand	Pentatricopeptide repeat-containing protein	AT4G16390.1
GRMZM2G128877	5	4,738	6,711	180	542	Dregion	Vacuolar ATP synthase subunit A	AT1G78900.1
GRMZM2G130055	5	774	288	21	14	DI_region	"Antifreeze protein, type I"	
GRMZM2G134563	4	178	367	10	25	Dregion	Tetratricopeptide repeat-like superfamily protein	AT1G77230.1
GRMZM2G135970	6	7,197	5,053	9	3	Dregion	Peptidase M20/M25/M40 family protein	AT1G44820.1
GRMZM2G137947	1	8,441	6,701	67	20	Dregion		
GRMZM2G155015	6	2,214	3,198	0	5	Iregion	EamA-like transporter family	AT3G07080.1
GRMZM2G174092	2	271	467	69	30	Iregion		
GRMZM2G175141	1	2,923	1,005	2	12	Dcand	Uncharacterized	
GRMZM2G177620	10	434	96	7	7	Iregion		
GRMZM2G339488	1	10,620	7,395	8	7	Icand		
GRMZM2G351023	2	464	1,248	3	37	Iregion	NAD(P)-binding Rossmann-fold superfamily protein	AT1G24360.1
GRMZM2G361633	6	4,379	7,331	11	186	Icand		
GRMZM2G371721	7	5,089	3,551	26	97	Dregion	Chaperone DnaJ-domain superfamily protein	AT5G64360.3
GRMZM2G381051	6	7,473	5,661	50	16	Dregion	Isovaleryl-CoA-dehydrogenase	AT3G45300.1
GRMZM2G406798	2	308	813	2	13	Iregion		
GRMZM2G444874	7	445	850	1	6	DI_region	2-Oxoglutarate and Fe(II)-dependent oxygenase superfamily protein	AT3G12940.1
GRMZM2G455122	2	798	194	0	4	Iregion		
GRMZM2G457211	5	255	413	1	4	Iregion	Protein of unknown function, DUF547	AT3G13000.2
GRMZM2G457231	5	913	1,706	48	34	Dregion	CTC-interacting domain 10	AT3G49390.1
GRMZM2G702552	4	172	586	12	94	Dregion	Uncharacterized	

avg, average; Chr, chromosome; Dom, domestication; Imp, improvement.

Other Supporting Information Files

Dataset S1 (XLSX)