Evidence for a natural allelic series at the maize domestication locus

*teosinte branched1*

Anthony J. Studer and John F. Doebley

*Department of Genetics, University of Wisconsin-Madison, Madison, Wisconsin 53706*
Short Title: Natural allelic series at tb1

Key Words: maize, teosinte, QTL, multiple alleles

Corresponding Author: Anthony Studer, Department of Genetics, 425 Henry Mall

Genetics/Biotech Building #5210, Madison WI 53706, 608-265-5804, tony.studer@gmail.com
ABSTRACT

Despite numerous quantitative trait loci and association mapping studies, our understanding of the extent to which natural allelic series contribute to the variation for complex traits is limited. In this study, we investigate the occurrence of a natural allelic series for complex traits at the *teosinte branched1* (*tb1*) gene in natural populations of teosinte (*Zea mays* ssp. *parviglumis*, *Z. mays* ssp. *mexicana*, and *Z. diploperennis*). Previously, *tb1* was shown to confer large effects on both plant architecture and ear morphology between domesticated maize and teosinte; however, the effect of *tb1* on trait variation in natural populations of teosinte has not been investigated. We compare the effects of nine teosinte alleles of *tb1* that were introgressed into an isogenic maize inbred background. Our results provide evidence for a natural allelic series at *tb1* for several complex morphological traits. The teosinte introgressions separate into three distinct phenotypic classes, which correspond to the taxonomic origin of the alleles. The effects of the three allelic classes also correspond to known morphological differences between the teosinte taxa. Our results suggest that *tb1* contributed to the morphological diversification of teosinte taxa as well as to the domestication of maize.
INTRODUCTION

Over the past several decades, there has been considerable interest in the genetic architecture of trait variation in natural populations as defined by number of genes involved and the effect sizes of these genes (Tanksley 1993; Mackay 2001). A key component of this issue is how variation is structured at individual genes. Are genes typically biallelic, like Mendel’s classic loci, or do genes often harbor allelic series, i.e. multiple alleles with measurably different effects on traits? While allelic series are known for pigmentation and other simple phenotypic traits, such as the extension allelic series controlling coat color in rabbits (Fontanesi et al. 2006), allelic series for complex morphological traits are not well-documented. The unambiguous documentation of a natural allelic series for complex traits would further the understanding of the genetic architecture of variation in natural populations.

Maize and its wild relatives, the teosintes, are an attractive system for the study of natural variation and complex traits. Maize and the teosintes belong to the genus Zea, which has four species that are native to Mexico and Central America: Z. perennis, Z. luxurians, Z. diploperennis and Z. mays (Doebley and Iltis 1980). The latter species includes four subspecies: one for domesticated maize (ssp. mays) plus three subspecies for teosinte (sspp. parviglumis, mexicana and huehuetenangensis), each with a distinct eco-geographic distribution. Of these three wild subspecies, ssp. parviglumis has been identified as the wild progenitor of maize (Doebley 2004). Since these teosinte taxa are interfertile with maize, one can leverage the genetic tools of maize to study variation in teosinte. Some of these teosinte taxa are widespread and contain abundant natural genetic variation (Fukunaga et al. 2005). The teosintes are an appealing gene pool in which one could search for natural allelic series for complex traits.
Among the ~35,000 maize genes, an attractive candidate for the study of natural allelic series is *teosinte branched1* (*tb1*). This gene controls plant architecture (apical dominance) and ear morphology (Doebley *et al.* 1997). *tb1* is a member of the TCP family of transcription factors (Cubas *et al.* 1999), and it is one of the key genes involved in the domestication of maize (Doebley 2004). During maize domestication, ancient farmers selected an allele of *tb1* that is expressed about twice as strongly as most teosinte alleles. The factor controlling this difference in gene expression has been mapped to a regulatory region 58 to 69 kb upstream of the *tb1* ORF (Clark *et al.* 2006; Studer *et al.* 2011). Since teosinte possessed natural allelic variation at *tb1* upon which ancient farmers could apply selection, it seems plausible that teosinte might contain a natural allelic series at this gene for traits related to plant architecture and ear morphology.

In this paper, we present evidence for a natural allelic series at *tb1*. We introgressed nine teosinte chromosomal segments encompassing *tb1* into the isogenic background of a maize inbred line. These *tb1* containing segments included four from *Z. mays* ssp. *mexicana*, four *Z. mays* ssp. *parviglumis* and one *Z. diploperennis*. We compare the effects of these *tb1* introgressions to one another and to a maize reference allele for four morphological traits that are known to be controlled by this gene (Clark *et al.* 2006; Studer *et al.* 2011). We show that the introgressed teosinte *tb1* chromosomal segments separate into three distinct phenotypic classes and that these classes correspond to the taxonomic origin of the segments. Moreover, the effects of the *tb1* segments match the known morphological differences between these taxa. Our results suggest that *tb1*, which contributed to maize domestication, also played a role in the morphological divergence of teosinte taxa.
MATERIALS AND METHODS

**Plant materials**: Segments of the long arm of chromosome 1 from nine different teosintes (IS1-9, Table S1) were introgressed into a maize inbred W22 background via six generations of backcrossing. The nine teosintes were chosen to represent multiple taxa over a wide geographic range without a knowledge of the sequence makeup at \(tb1\). During the backcrossing process, RFLP markers (NPI615, umc140, \(tb1\), umc107, bnl15.18, \(kn1\)) flanking \(tb1\) were used to follow the target segment. After the 6\(^{th}\) generation of backcrossing, the BC\(_6\) plants were selfed and PCR-based markers were used to map each of the teosinte introgressed chromosomal segments (Figure 1, Table S2).

**Phenotypic data collection and analysis**: Plants were grown at the University of Wisconsin West Madison Agricultural Research Station, Madison, WI, USA during the summer of 2009. For each of the nine introgressed teosinte \(tb1\) chromosomal segment, a population of 140 BC\(_6\)S\(_1\) plants was grown. All 1260 plants were grown together in a single fully randomized plot with 0.9 meter spacing between plants in both dimensions. This spacing minimized the degree to which plants shaded their neighbors. Seedlings were genotyped using two PCR-based indel markers in \(tb1\) (Table S2). Phenotypic analysis was performed on all plants that were homozygous for either the maize or teosinte marker genotype. Using BC\(_6\)S\(_1\) plants allowed us to compare individuals containing the introgressed teosinte chromosomal segments with individuals homozygous for the W22 segment. Seed for each of the nine populations was obtained from a single ear, thus eliminating any concern that differences among genotypic classes within a population are due to ear-parent effects.

The following four traits were phenotyped: cupules per rank (CUPR; number of cupules in a single rank from base to the tip of the ear), lateral branch internode length (LBIL; mean
internode length, in cm, for the uppermost lateral branch), tillering (TILL; sum of tiller heights/plant height), and percent staminate spikelets (STAM; percentage of male spikelets in the inflorescence). CUPR and STAM were both measured on the uppermost lateral inflorescence (ear) of each plant.

The software package JMP IN 4.0.4 (SAS Institute Inc., Cary, NC) was used for calculating means, standard errors, Levene's tests, and principal component analysis (PCA). The MIXED procedure of SAS (SAS Institute Inc., Cary, NC) was used to implement a mixed linear statistical model to test for genotypic, family, and allelic effects. Genotype (W22 control segment vs. teosinte introgression at \(tb1\)) was considered a fixed effect, while family (IS1-IS9) and a genotype by family interaction term were treated as random effects. The full linear model used was

\[
Y_{cdf} = \mu + a_c + b_d + a_c*b_d + e_{cdf}
\]

where \(Y_{cdf}\) is the trait value for the \(f^{th}\) plant with \(c^{th}\) genotype from the \(d^{th}\) family, \(\mu\) is the overall mean of the experiment, \(a_c\) is the genotypic effect, \(b_d\) is the family effect, \(a_c*b_d\) is the genotype by family interaction, and \(e_{cdf}\) is the error term. Terms were added individually to the model and tested for significance using the Likelihood Ratio Test which has a Chi-squared distribution with one degree of freedom. Levene’s test was used to assess the equality of the variance for plants containing the W22 control segments vs. plants containing teosinte introgression segments. PCA was performed using the trait correlation matrix for the additive effects. Additive effects were calculated as half the difference between the mean of the homozygous teosinte introgression class and the mean of the corresponding maize class. The sample size for each introgression family and the additive effect estimates are listed in Table 1. Given that the difference between
W22 and itself equals zero, zero was used for all W22 trait values in the PCA. All genotype and phenotype data are available at www.panzea.org.

**Nucleotide sequence analysis:** Polymerase chain reaction (PCR) was done using Qiagen Taq DNA Polymerase following the manufactures instructions and standard methods. One primer set was used to amplify the \( tb1 \) coding region (GGACATATGAGTAGGCCACACTCCTCC, GATTTGCAGCTCATCAAGAAA) and two additional internal primers were used to sequence the PCR product (TCATGGACAACGATGAGTGG, CCAAGAAAATCGGCCAATAA). Two primer sets were used to amplify the \( tb1 \) control region (CGGTCAAAGAGTAGGGCAAG, CGTGTGTGATCGAATGGT). Sequencing of PCR fragments was done using Applied Biosystems (ABI) BigDye and an ABI 3730xl DNA Analyzers at the University of Wisconsin Biotechnology Center DNA Sequencing Facility. Initial alignment of nucleotide sequences was performed using ClustalW (Thompson *et al.* 1994) and then finished by hand using MEGA version 5.03 (Tamura *et al.* 2011). Neighbor Joining Trees were constructed in PAUP 4.0b10 (Swofford 2003) using the absolute number of differences after gaps, missing and ambiguous bases were removed from the alignment.
RESULTS

**Genotypic, family, and allelic effects:** To observe each of the *tb1* teosinte chromosomal segments’ effect on phenotype relative to the control W22 chromosome segment, the mean for plants homozygous for each of the nine introgression segments and the mean for each of the control populations (plants homozygous for the maize segment) were plotted for the four phenotypes (Figure 2): the number of cupules per rank (CUPR), the internode length on the uppermost lateral branch (LBIL), the amount of tillering (TILL), and the percent staminate spikelets on the primary lateral inflorescence (STAM). These traits represent ear morphology (CUPR and STAM) and plant architecture (LBIL and TILL), which are some of the major morphological differences between maize and teosinte. One of the teosinte introgressed segments (IS6) does not have data for ear morphology traits because all of the ears (from introgression and control plants) were sterile (without kernels).

To test for genotypic (W22 control segment vs. teosinte introgression), family (IS1-IS9), and allelic affects (genotype*family) a mixed linear statistical model was used. For all traits, the model indicates that plants containing a teosinte introgression segment are significantly different from plants with a W22 control segment (Table 2A, Figure 2). A significant family effect is also supported by the model for each trait; indicting a difference between family means (Table 2A). To explore the differences between families further, an interaction term was added to the model to ask if the introgression family affects both genotypes equally. If significant, this interaction would suggest that not all of the teosinte introgression segments are equivalent, assuming that all of the W22 controls engender equivalent phenotypes given they carry the same allele at *tb1*. The interaction term is significant for LBIL, CUPR and STAM but not for TILL (Table 2A). The insignificant interaction term for TILL indicates that all of the teosinte introgressed segments
have equivalent effects, and that the variance observed among families is due to factors other
than an allelic series at tb1.

We observed a significant family effect for TILL but not a significant family by genotype
interaction term. This result suggests that there are differences among the introgression families
due to a factor other than the tb1 introgression segment. Two possible explanations were
considered for this result. First, significant phenotypic differences between families could be
observed if additional genetic factors segregated between backcrossing populations at loci in the
genome unlinked to the target segment encompassing tb1. Such factors would increase (or
decrease) the trait mean for both plants with the introgressed teosinte chromosomal segment and
the corresponding control plants, which would contribute to a significant family term in the
model but not a significant interaction between genotype and family. Second, environmentally
determined seed quality differences among the ear-parents for different introgression families
could be responsible. This is particularly possible since only a single ear parent was used for
each introgression families. Ear-parent effects such as seed weight, seed maturity and speed of
germination can influence adult phenotype. Thus, environmentally induced ear-parent effects
could account for the differences seen among the introgression families, which were derived
from different ears.

For LBIL, the interaction term between genotype by family was found to be significant
(Table 2A), suggesting that not all of the teosinte introgression segments are equivalent when
assuming all of the W22 control populations are equivalent. To investigate this result further, a
model with just a family term was tested against a null model. This was performed separately for
the W22 control and the teosinte introgression subsets of the data. The family term was
significant for both subsets (Table 2B), indicating that while there are significant differences
among the teosinte introgressions this could be the result of factors other than an allelic series at $tb1$ since there are also significant differences among the W22 controls.

Since there were significant differences among W22 controls for LBIL, we used the Levene's test to ask whether the variance among teosinte $tb1$ introgressions for LBIL is equivalent to the variance among the control populations as expected if there is not an allelic series at $tb1$. The results of this test indicate that there is greater variance among teosinte introgressions as compared to the control populations (Table 3), suggesting that the teosinte introgressions possess different allelic effects for LBIL. A graph of the additive effects for LBIL highlights the small effect of IS5, and the large effect IS7 has on LBIL compared to the rest of the teosinte introgressions (Figure 3).

Our analyses using the mixed linear model for the ear morphology traits (CUPR and STAM) produced a different result than seen for the plant architecture traits (TILL, LBIL). For both CUPR and STAM, the interaction term for genotype by family was found to be significant (Table 2A), suggesting that not all of the teosinte introgression segments are equivalent. Furthermore, when the interaction was included in the model, the family term dropped out, indicating that all of the variance observed among families is due to differences between teosinte introgression segments rather than unlinked genetic factors that differ among the introgression populations or ear-parent effects as discussed above. A model with just a family term was used to test the assumption that the W22 control populations are equivalent. Significant family effects were found among teosinte introgression segments but not among W22 controls (Table 2B), further supporting the hypothesis that there is an allelic series for CUPR and STAM at $tb1$.

The $tb1$ introgressions form distinct classes: To assess whether the different teosinte $tb1$ introgressions could be classified into groups based on phenotype, a principal component
analysis was performed using the additive effects of the four traits as input data. IS6 was not included in the analysis because it is missing data for two of the four traits. Two components were retained from the analysis, which explain 64% and 27% of the observed variance. The ear morphology traits, cupules per rank and staminate spikelets, load to component 1, which is represented by the x-axis in Figure 4. The plant architecture traits, tillering and lateral branch internode length, load to component 2, which is represented by the y-axis in Figure 4. The W22 control plots to the lower left quadrant of the graph with distance from this point corresponding to more teosinte-like phenotypes (Figure 4).

The principal component analysis suggests that there are three classes of teosinte \textit{tb1} introgressions. The first class is composed of a single introgression (IS5), which plots away from the rest of the teosinte introgressions and is located in the quadrant containing the W22 control. This result suggests that IS5 is an allele that confers a phenotype which is only modestly different from the W22 control. This relationship can also be observed by looking at IS5 for each trait individually (Figures 2-3). The second class is composed of IS2, 4, 7, and 9, all of which plot to the upper left quadrant (Figure 4). This quadrant represents introgressions that produce teosinte-like plant architecture traits (long tillers and lateral branches), but maize-like ear morphology traits (more cupules per rank and few staminate spikelets). The final class is composed of IS1, 3, and 8 and occupies the right half of the graph along the x-axis (Figure 4). These introgressions produce both a more teosinte-like ear morphology and plant architecture. In particular, IS1, 3, and 8 have a high percentage of male spikelets in their ears (Figures 2-3).

Strikingly, the PCA reveals that the allelic classes correspond to the taxonomic origin of the teosinte \textit{tb1} introgression (Figure 4). The allelic class with the most teosinte-like phenotypes corresponds to introgressions from \textit{Z. mays} ssp. \textit{parviglumis} (PAR). The allelic class with
moderate teosinte-like phenotypes corresponds to introgressions from the *Z. mays* ssp. *mexicana* (MEX). Finally, the allelic class with the most maize-like phenotypes corresponds to the introgression from *Z. diploperennis* (DIP). Thus, the allelic series at *tb1* appears to have a taxonomic basis. Because of the isogenic nature of the introgression lines, the apparent allelic series can not the result of factors other than a difference at or near *tb1*.

Although the allele series shows a distinct taxonomic signature, we also asked whether the allele classes were correlated with the length of the introgressed segments (Figure 1). No obvious correlation between phenotype and introgression length is observed. For example, the largest introgression (IS2) does not have the most teosinte-like phenotypes, nor does the smallest introgression (IS9) have the most maize-like phenotypes (Figure 3). Moreover, different introgression lengths are represented in the different allelic classes defined in the PCA. This result supports the conclusion of an allelic series at *tb1*, as opposed to other linked genes in the introgressed segments causing the observed allelic differences.

To explore the possibility of a correlation between the nucleotide sequence of *tb1* and phenotype, we plotted the phenotypic classes defined by the PCA onto neighbor-joining trees based upon two regions of the *tb1* nucleotide sequence (Figure 5). One portion is the protein coding region of the gene and 3’ UTR, and the other corresponds to a known upstream regulatory region of *tb1* (Clark *et al.* 2006). The teosinte introgressions representing any single allelic class defined by the PCA are scattered across both of the phylogenetic trees, and for the most part, no relationship between phylogeny and phenotype is apparent. For example, the class representing the most teosinte-like ear phenotypes (IS1, 3, and 8) does not cluster in either phylogeny. One striking feature of both phylogenies is that IS5, which was derived from a separate species (*Z. diploperennis*, DIP) and has unique phenotypic effects, stands apart from all
other introgressions in both trees. This result suggests that the different phenotypes observed for
IS5 when compared to the other introgressions could be due to sequence differences in the
upstream control region and/or the coding sequence of \( tb1 \). Since the introgressions from neither
\( Z. \ mays \ ssp. \ parviglumis \) (PAR) nor \( Z. \ mays \ ssp. \ mexicana \) (MEX) cluster on either of the
phylogenetic trees, these trees do not enable to identify sequence variants that control the
putative allelic variation.

DISCUSSION

Natural allelic series for simple phenotypic traits such as pigmentation are well
documented in the literature. For example, five alleles have been described at the \( R \) locus in
maize, which control plant and kernel pigmentation. Each of these 5 alleles produces a distinct
phenotype based on pigment quantity, spatial patterning in kernels, the timing of pigmentation
onset during development, and which organs are pigmented (kernels, anthers, leaves and/or roots)
(Styles et al. 1973). A similar allelic series for pigmentation has been described for the \( B \) locus
of maize (Styles et al. 1973; Radicella et al. 1992). Much like these examples from maize, an
allelic series for coat color in mice has been described (Phillips 1966; Jackson 1994). Alleles of
the \textit{agouti} locus produce distinct coat colors and pattern differences due to factors in both the
promoter and coding region of the gene. Allelic series have also been described for traits such as
self-incompatibility in plants (Nasralla et al. 1991; Takayama and Isogai 2005).

Evidence for natural allelic series for complex or morphological traits has come from
association mapping and QTL studies (Purugganan and Suddith 1998; Todesco et al. 2010;
McKechnie et al. 2010). For example, an allelic series for flowering time was reported among a
diverse set of maize lines which display significant variation in flowering time (Buckler et al.
In this example, statistical evidence for an allelic series is shown; however there is no actual proof that a single locus with multiple alleles explains the observed phenotypic series since the occurrence of several tightly linked genes each with two alleles cannot be excluded. Another concern with evidence for allelic series from QTL and association studies is that the alleles are each typically characterized in a different genetic background. Thus, it is possible that the QTL in question has only two alleles which form a number of apparent allelic classes based on the background in which they were assayed. Using association mapping, Weber et al. (2007) assayed variation in a natural teosinte population and found multiple SNPs in and around tb1 associated with small effects on plant and ear architecture. Their results are consistent with our data, which show relatively small amounts of natural variation within taxa. However, because Weber et al. (2007) only included a single taxon (Z. mays ssp. parviglumis), the among taxa variation that we described may be distinct from the allelic effects they report.

In this paper, we present evidence for a natural allelic series at tb1 for three complex morphological traits: lateral branch internode length, the number of cupules per rank, and the number of staminate spikelets (Figures 2-3). Our evidence for allelic series at tb1 largely eliminates concerns about the influence of genetic background by using isogenic lines. We also examined the role of linked genes on trait variation associated with tb1 by considering the length of the introgressed chromosomal segment surrounding tb1 for each of the teosinte introgressions. We saw no evidence that phenotype is correlated with the length of the introgression segment (Figures 1, 4).

An argument could be made that tb1 is contributing little or nothing to the observed phenotypic variation that we observed, and that the variation is caused by heterogeneity at linked genes, given that the introgressions contain between roughly 80 (IS9) and more than 800 (IS2)
linked genes. However, IS9, which spans only 80 linked genes, has nearly identical phenotypic effects to IS2, which spans 800 linked genes, arguing against a role for linked genes contributing to the observed phenotypic variation. Ideally, teosinte introgression segments of a uniform length and which only contain the \textit{tb1} gene itself would be compared. However, the creation of such lines would be a long and labor intensive process.

A recent report (Studer and Doebley 2011) on the fractionation of the QTL effects at \textit{tb1} sheds additional light on whether linked genes underlie the allelic effects that we observed. This report shows that the QTL at \textit{tb1} for plant architecture traits (TILL and LBIL) does not fractionate, but rather maps narrowly to a 69 kb region upstream of the \textit{tb1} coding sequence. This result is consistent with the hypothesis that the variation for LBIL near \textit{tb1} is due sole to \textit{tb1}, and by inference, that the variation observed in this study is attributable to allelic differences at \textit{tb1} and not linked genes.

Studer and Doebley (2011) report that the QTL at \textit{tb1} for CUPR does fractionate, however, presence/absence of the teosinte allele at this QTL does not correlate well with the phenotypic differences among the introgressions. For example, IS1, 2, 3 and 8 all carry the teosinte allele of the CUPR QTL identified upstream of \textit{tb1}. While IS1, 3 and 8 all show large effects for CUPR, IS2 has only a small effect (Figure 3). Furthermore, \textit{Z. mays} ssp. \textit{mexicana} introgression IS7 has an intermediate effect on CUPR but does not have the teosinte allele for the CUPR QTL. The taxonomic origin of the introgression is a better predictor of variation for CUPR than the presence/absence of the teosinte allele for the CUPR QTL. While \textit{Z. mays} ssp. \textit{parviglumis} introgressions (IS1, 3, and 8) all have strong effects, most \textit{Z. diploperennis} and \textit{Z. mays} ssp. \textit{mexicana} introgressions (IS2, 4, 5 and 9) have weak effects (Figure 3).
Studer and Doebley (2011) also report that the QTL at \(tb1\) for STAM fractionates, and in this case, there is a correlation, although imperfect, between presence/absence of the teosinte allele at this QTL and the phenotypic differences among the introgressions for STAM. Our introgressions IS1, 3, and 8 all carry the teosinte allele of this QTL for STAM and have the largest effects for STAM, while IS4, 5, 7 and 9 carry the maize allele and show no effect on STAM (Figure 3). The exception is IS2 which carries the teosinte allele at this linked STAM QTL but does not have a significant effect on STAM. Thus, taxonomy is still a better predictor of variation for STAM than the presence/absence of the teosinte allele for the STAM QTL since \(Z.\ mays\ ssp.\ parviglumis\) introgressions (IS1, 3, and 8) all have strong effects and \(Z.\ diploperennis\) and \(Z.\ mays\ ssp.\ mexicana\) introgressions (IS2, 4, 5, 7 and 9) all have weak effects (Figure 3).

The feature that is best correlated with the phenotypic effects of the \(tb1\) alleles that we examined is the taxonomic origin of these alleles. In a principal components analysis based on phenotype, the eight teosinte introgressions form three classes that correspond to \(Z.\ mays\ ssp.\ parviglumis\), \(Z.\ mays\ ssp.\ mexicana\) and \(Z.\ diploperennis\) (Figure 4). This result not only supports the existence of an allelic series at \(tb1\), but it also implicates \(tb1\) in the morphological diversification of these taxa in addition to its role in maize domestication. There are several notable correspondences between known morphological differences between these taxa and the effects associated with the alleles of \(tb1\) we assayed. First, \(Z.\ mays\ ssp.\ mexicana\) has more fruitcases (a greater CUPR value) per ear than either \(Z.\ mays\ ssp.\ parviglumis\) or \(Z.\ diploperennis\) (Iltis and Doebley 1980), and our \(Z.\ mays\ ssp.\ mexicana\) alleles have greater CUPR values than our \(Z.\ mays\ ssp.\ parviglumis\) and \(Z.\ diploperennis\) alleles (Figure 3). Second, \(Z.\ diploperennis\) has shorter lateral branches that are tipped in a mixed male-female inflorescence.
unlike other teosintes that have longer lateral branches tipped by tassels (Iltis et al. 1979; Doebley and Iltis 1980). The one Z. diploperennis allele we assayed has the smallest value for LBIL (shorter branches) of all nine teosinte alleles assayed (Figure 3). While our observations suggest that tb1 may partly control morphological differences among teosinte taxa, our study includes a limited sampling of each taxa and thus the data must be regarded as suggestive rather than conclusive.

Given the correlation between taxonomy and allelic effects (Figure 4), we examined phylogenetic trees based on the nucleotide sequences of the control region and coding sequence of tb1, but we saw no relationship between phenotype and phylogeny. We also examined the sequence alignments for any fixed differences between the taxa that may not have been visible in the trees. No fixed differences were found between Z. mays ssp. mexicana and Z. mays ssp. parviglumis individuals for either sequenced region. The Z. diploperennis sequence is highly divergent from the other alleles with many sequence differences. With such a large number of differences and only a single Z. diploperennis sample, it is not possible to say which if any are potentially causative. However, there are two polymorphisms unique to the Z. diploperennis allele of tb1 that cause radical amino acid changes in the Helix II portion of the TCP domain, which is involved in DNA binding. An A>G substitution at AGP_v2 position 265746492 causes a T to A amino acid change, and T>G substitution at AGP_v2 position 265746501 causes a S to A amino acid change. Both changes are from hydrophobic to hydrophilic amino acids, which could alter protein function. Further experimentation is needed to test whether these amino acid differences affect phenotype.

In summary, our experiments provide evidence for a natural allelic series at tb1 with effects on complex morphological traits. It has been previously shown that tb1 played a major
role in the domestication of maize from its wild progenitor, teosinte (Doebley 2004). Since the allelic classes that we observed at tb1 correspond with taxonomic origin, tb1 may also have played a role in the morphological diversification of Z. mays ssp. parviglumis, Z. mays ssp. mexicana and Z. diploperennis. To provide final proof of the allelic series at tb1 and verify its role in the divergence of teosinte, the causal polymorphisms underlying the phenotypic differences needs to be identified.

We thank Brian Yandell and Peter Bradbury for statistical advice concerning the data analyses in this manuscript. This work was supported by the USDA Hatch grant MSN101593 and the National Science Foundation grants DBI0321467 and DBI0820619.

Sequence data from this article has been deposited with Genbank as accessions JQ900488-JQ900509.
LITERATURE CITED


TABLE 1

N and additive effects for homozygous plants

<table>
<thead>
<tr>
<th>Introgression</th>
<th>$N_a / N_T$</th>
<th>Additive Effects</th>
<th>Additive Effects</th>
<th>Additive Effects</th>
<th>Additive Effects</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>TILL</td>
<td>LBIL</td>
<td>CUPR</td>
<td>STAM</td>
</tr>
<tr>
<td>IS1</td>
<td>21 / 32</td>
<td>0.6993</td>
<td>0.8075</td>
<td>-3.2368</td>
<td>2.9186</td>
</tr>
<tr>
<td>IS2</td>
<td>34 / 31</td>
<td>0.5684</td>
<td>0.7766</td>
<td>-0.4941</td>
<td>0.3971</td>
</tr>
<tr>
<td>IS3</td>
<td>23 / 28</td>
<td>0.6558</td>
<td>0.5455</td>
<td>-3.3727</td>
<td>2.1882</td>
</tr>
<tr>
<td>IS4</td>
<td>33 / 41</td>
<td>0.5619</td>
<td>0.8682</td>
<td>-0.2175</td>
<td>-0.0095</td>
</tr>
<tr>
<td>IS5</td>
<td>33 / 23</td>
<td>0.4446</td>
<td>-0.0592</td>
<td>-0.6667</td>
<td>0.0000</td>
</tr>
<tr>
<td>IS6</td>
<td>31 / 34</td>
<td>0.3539</td>
<td>0.7284</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>IS7</td>
<td>32 / 29</td>
<td>0.6757</td>
<td>1.4879</td>
<td>-1.9267</td>
<td>0.0000</td>
</tr>
<tr>
<td>IS8</td>
<td>37 / 30</td>
<td>0.5191</td>
<td>0.9026</td>
<td>-3.0700</td>
<td>2.2961</td>
</tr>
<tr>
<td>IS9</td>
<td>24 / 32</td>
<td>0.6370</td>
<td>0.5928</td>
<td>-0.8644</td>
<td>0.0273</td>
</tr>
</tbody>
</table>

*a“M” denotes homozygous maize control plants, “T” denotes homozygous teosinte introgression plants.
TABLE 2

Likelihood ratio test results
comparing statistical models with genotype, family, and allelic effects (d.f. = 1)

<table>
<thead>
<tr>
<th></th>
<th>Genotype&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Family&lt;sup&gt;b&lt;/sup&gt;</th>
<th>Genotype*Family&lt;sup&gt;c&lt;/sup&gt;</th>
<th>Allelic Effects</th>
</tr>
</thead>
<tbody>
<tr>
<td>A.</td>
<td>Test</td>
<td>M vs. T</td>
<td>IS1… IS9</td>
<td>0.1573</td>
</tr>
<tr>
<td></td>
<td>TILL</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
<td></td>
</tr>
<tr>
<td></td>
<td>LBIL</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td></td>
<td>CUPR</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td></td>
<td>STAM</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>B.</td>
<td>Family (M)&lt;sup&gt;d&lt;/sup&gt;</td>
<td>Family (T)&lt;sup&gt;d&lt;/sup&gt;</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>TILL</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
<td></td>
</tr>
<tr>
<td></td>
<td>LBIL</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
<td></td>
</tr>
<tr>
<td></td>
<td>CUPR</td>
<td>0.3711</td>
<td>&lt;0.0001</td>
<td></td>
</tr>
<tr>
<td></td>
<td>STAM</td>
<td>1</td>
<td>&lt;0.0001</td>
<td></td>
</tr>
</tbody>
</table>

<sup>a</sup> Tests if the means of both genotypes are equal, or if the two genotypic means are different.

\[ H_0: Y = \mu + e, H_a: Y = \mu + \text{Genotype} + e \]

<sup>b</sup> Tests if the nine family means are equal, or if two or more family means are not equal.

\[ H_0: Y = \mu + \text{Genotype} + e, H_a: Y = \mu + \text{Genotype} + \text{Family} + e \]

<sup>c</sup> Tests if the allelic effects between families are equal, or if two or more families have allelic effects that are not equal.

\[ H_0: Y = \mu + \text{Genotype} + \text{Family} + e, H_a: Y = \mu + \text{Genotype} + \text{Family} + \text{Genotype*Family} + e \]

<sup>d</sup> Tests if the nine family means are equal, or if two or more family means are not equal using maize (M) and teosinte introgression (T) subsets of the data.

\[ H_0: Y = \mu + e, H_a: Y = \mu + \text{Family} + e \]
<table>
<thead>
<tr>
<th>Test</th>
<th>Levene's M vs. T</th>
<th>Levene's DFNum</th>
<th>Levene's DFDen</th>
<th>Levene's F-ratio</th>
<th>Levene's P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>TILL</td>
<td>1</td>
<td>16</td>
<td></td>
<td>0.0089</td>
<td>0.9262</td>
</tr>
<tr>
<td>IBIL</td>
<td>1</td>
<td>16</td>
<td></td>
<td>5.0864</td>
<td>0.0385</td>
</tr>
<tr>
<td>CUPR</td>
<td>1</td>
<td>14</td>
<td></td>
<td>40.6558</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>STAM</td>
<td>1</td>
<td>14</td>
<td></td>
<td>62.6364</td>
<td>&lt;0.0001</td>
</tr>
</tbody>
</table>
FIGURE 1.-Physical map of the introgression lines. All introgressed segments are drawn to scale, and vertical dotted lines show AGPv2 reference position (Mb). Shaded areas indicate teosinte chromosome segments based on taxonomic origin: (blue) *Zea diploperennis*, (red) *Z. mays* ssp. *parviglumis*, and (green) *Z. mays* ssp. *mexicana*; unshaded areas represent maize chromosome segments. Markers used for genotyping are shown along the chromosomes as solid black lines and listed in Table S2. The position of *tb1* is shown for reference.

FIGURE 2.-Phenotypic means. Points are shaded based on taxonomic origin of the *tb1* introgressed segment: (purple) *Zea mays* ssp. *mays* control populations, (blue) *Z. diploperennis*, (red) *Z. mays* ssp. *parviglumis*, and (green) *Z. mays* ssp. *mexicana*. Error bars represent the standard error for each genotypic class. The x-axis shows the introgression segments; the y-axis shows trait means.

FIGURE 3.-Additive effects. Traits are abbreviated as follows: cupules per rank (CUPR), lateral branch internode length (LBIL, in cm), staminate spikelets (STAM, percent), and tillering (TILL). The x-axis shows the introgression segments; the y-axis shows additive effects. Error bars represent the standard error for each effect. Bars are shaded based on taxonomic origin of the introgression segments: (blue) *Zea diploperennis*, (red) *Z. mays* ssp. *parviglumis*, and (green) *Z. mays* ssp. *mexicana*.

FIGURE 4.-Principle components plot. The x-axis shows component 1 which represents ear morphology traits; the y-axis shows component 2 which represents plant morphology traits. Dots are shaded based on taxonomic origin of the introgression segment: (purple) *Zea mays* ssp. *mays*, (blue) *Z. diploperennis*, (red) *Z. mays* ssp. *parviglumis*, and (green) *Z. mays* ssp. *mexicana*.

FIGURE 5.-Phylogenetic trees. (A) A neighbor joining tree based on sequence from the *tb1* coding region. (B) A neighbor joining tree based on sequence from the *tb1* upstream control region. Text color is based on taxonomic origin of the introgression segment: (purple) *Zea mays* ssp. *mays*, (blue) *Z. diploperennis*, (red) *Z. mays* ssp. *parviglumis*, and (green) *Z. mays* ssp. *mexicana*. 
# TABLE S2

## Markers for genotyping

<table>
<thead>
<tr>
<th>Markers</th>
<th>Forward Primer (5’ to 3’)</th>
<th>Reverse Primer (5’ to 3’)</th>
</tr>
</thead>
<tbody>
<tr>
<td>umc2569*a</td>
<td>GTGACACCCCTAGCCCTTTAGCA</td>
<td>TAGCTTGGATATGTGGTTG</td>
</tr>
<tr>
<td>umc2237*a</td>
<td>CTGACTTACAGGACAGAAGG</td>
<td>GTCACTGACATATCCTCAGAGCC</td>
</tr>
<tr>
<td>umc1122*a</td>
<td>CACAATCTGACAGGAGAGAAGG</td>
<td>CTGACTGACATATCCTCAGAGCC</td>
</tr>
<tr>
<td>umc2396*a</td>
<td>TGCTATTTTAGGGTTG</td>
<td>TGCATGACATATCCTCAGAGCC</td>
</tr>
<tr>
<td>bnlg1615*a</td>
<td>CTTCCTTCGCCCTCTTGGTTCA</td>
<td>GCACTGACATATCCTCAGAGCC</td>
</tr>
<tr>
<td>bnlg1025*a</td>
<td>TGGTGAAGGGGAGAGATGAAG</td>
<td>CGAGACTGACTTCTAGGAAGG</td>
</tr>
<tr>
<td>bnlg1564*a</td>
<td>ACAGGACAGAAGAGAAGG</td>
<td>CTCTCCCTACATCCGCC</td>
</tr>
<tr>
<td>bnlg1629*a</td>
<td>GTCATGCATTTTAGGGTTG</td>
<td>TGGTGAAGGGGAGAGATGAAG</td>
</tr>
<tr>
<td>bnlg2228*a</td>
<td>GCACCAATCGACAGAGAAG</td>
<td>CTGACATCGACATCCGCC</td>
</tr>
<tr>
<td>umc2181*a</td>
<td>ATCGGGTCCGGTAGATTAGTAC</td>
<td>GTAGCTATAGCTAGCAGTGCGCC</td>
</tr>
<tr>
<td>mmc0041*a</td>
<td>AGGACTTACAGGAGAAGG</td>
<td>TTTATCCTACAGTCCGCC</td>
</tr>
<tr>
<td>umc1924*a</td>
<td>GGAATATGAGAATTCATTAGTG</td>
<td>CTAACAAACTTCTAGGCC</td>
</tr>
<tr>
<td>umc1991*a</td>
<td>GAATATGAGAATTCATTAGTG</td>
<td>CTAACAAACTTCTAGGCC</td>
</tr>
<tr>
<td>umc1914*a</td>
<td>CACAGAGAGAGAGAGGAG</td>
<td>TTTATCCTACAGTCCGCC</td>
</tr>
<tr>
<td>umc2047*a</td>
<td>GACAGACATTCGCTCTAGGCAC</td>
<td>CTGACATCGACATCCGCC</td>
</tr>
<tr>
<td>umc1298*a</td>
<td>AGCCTAAACAAAAGAGAAGAAGAG</td>
<td>TTTATCCTACAGTCCGCC</td>
</tr>
<tr>
<td>PZD00117.indel1*a</td>
<td>CCCGCCGCCCCGTCGCTAAGT</td>
<td>ATGCCGCGGCAAGCGACCG</td>
</tr>
<tr>
<td>umc1306*a</td>
<td>CCAAAAGACCAAGAGGTAGAAG</td>
<td>CTCAACGGTAAATGTAGTGAGC</td>
</tr>
<tr>
<td>bnlg1502*a</td>
<td>AGGGCGCTAGGCAATGACAGGA</td>
<td>AGAAGGGTAGGTACATCCGCC</td>
</tr>
<tr>
<td>umc1082*a</td>
<td>CCAGCACTGTGCAATAGGTTAGG</td>
<td>GCTCAGTACAGAGGTATAGT</td>
</tr>
<tr>
<td>PZD00101.indel1*b</td>
<td>ATCGACAAACACATCTTCCG</td>
<td>GCTTGGACATGGTTATGCTGG</td>
</tr>
<tr>
<td>umc1726*a</td>
<td>GATGCCGAAAGAAAAGAGGAAAA</td>
<td>AGACTGAACCTAATCTTAAATGCG</td>
</tr>
<tr>
<td>bnlg1671*a</td>
<td>TCACGGTAGCAAGAGAAGG</td>
<td>CCAACACAGGATAGGTCGAGTC</td>
</tr>
<tr>
<td>umc1774*a</td>
<td>ATGGGAGATTGTTGTATATTGA</td>
<td>TACACGACAGTACAGAGAGAG</td>
</tr>
<tr>
<td>umc2223*a</td>
<td>ACTTCTGCAGAGGCAAGG</td>
<td>TTTGGAGACTGAAAGAAGAGATCG</td>
</tr>
<tr>
<td>umc1500*a</td>
<td>TCTCTGGTGGCTACAGAAG</td>
<td>TTTGGAGACTGAAAGAAGAGATCG</td>
</tr>
<tr>
<td>umc1421*a</td>
<td>TGCTACAGAGATGGATACACAT</td>
<td>AGTGGTGAAATGTGCCTAGGAAAT</td>
</tr>
<tr>
<td>GS1*b</td>
<td>AGGAGGACAGAAGGGAAAGG</td>
<td>GCTTCAGTACAGAGAGAGAG</td>
</tr>
<tr>
<td>CR Indel*c</td>
<td>CGGTCAGAAGAGATGGGCAAG</td>
<td>GCCTGCTAGTCCGCCAGTCCAG</td>
</tr>
</tbody>
</table>

*a Markers used to map the introgressed teosinte segments

*b Directly labeled FAM genescan marker used to genotype IS3 F2 population.

*c Agarose gel marker used to genotype all IS F2 populations except IS3.
<table>
<thead>
<tr>
<th>Line</th>
<th>Species</th>
<th>Subspecies</th>
<th>Country</th>
<th>State/Province</th>
<th>Population</th>
<th>Collector</th>
<th>Collection</th>
<th>Lat Deg</th>
<th>Lat Min</th>
<th>Long Deg</th>
<th>Long Min</th>
</tr>
</thead>
<tbody>
<tr>
<td>IS1</td>
<td><em>mays</em></td>
<td><em>parviglumis</em></td>
<td>Mexico</td>
<td>Guerrero</td>
<td>1 mile S of Palo Blanco</td>
<td>Beadle &amp; Kato</td>
<td>Site 4</td>
<td>17</td>
<td>25</td>
<td>-99</td>
<td>30</td>
</tr>
<tr>
<td>IS2</td>
<td><em>mays</em></td>
<td><em>mexicana</em></td>
<td>Mexico</td>
<td>Mexico</td>
<td>km 43 on hwy from Chalco to Amecameca</td>
<td>Iltis <em>et al.</em></td>
<td>28622</td>
<td>19</td>
<td>6</td>
<td>-98</td>
<td>42</td>
</tr>
<tr>
<td>IS3</td>
<td><em>mays</em></td>
<td><em>parviglumis</em></td>
<td>Mexico</td>
<td>Guerrero</td>
<td>30 km S of Chilpancingo</td>
<td>Beadle &amp; Kato</td>
<td>Site 2-3</td>
<td>17</td>
<td>12</td>
<td>-99</td>
<td>30</td>
</tr>
<tr>
<td>IS4</td>
<td><em>mays</em></td>
<td><em>mexicana</em></td>
<td>Mexico</td>
<td>Jalisco</td>
<td>10 km S of Degollado</td>
<td>M. Puga</td>
<td>11066</td>
<td>20</td>
<td>22</td>
<td>-102</td>
<td>11</td>
</tr>
<tr>
<td>IS5</td>
<td><em>diploperennis</em></td>
<td></td>
<td>Mexico</td>
<td>Jalisco</td>
<td>Zarza Mora, 2 km E of Las Joyas</td>
<td>Iltis <em>et al.</em></td>
<td>1250</td>
<td>19</td>
<td>35</td>
<td>-104</td>
<td>16</td>
</tr>
<tr>
<td>IS6</td>
<td><em>mays</em></td>
<td><em>parviglumis</em></td>
<td>Mexico</td>
<td>Guerrero</td>
<td>1 km N of Mazatlan</td>
<td>Beadle &amp; Kato</td>
<td>Site 1</td>
<td>17</td>
<td>30</td>
<td>-99</td>
<td>30</td>
</tr>
<tr>
<td>IS7</td>
<td><em>mays</em></td>
<td><em>mexicana</em></td>
<td>Mexico</td>
<td>Chihuahua</td>
<td>Nobogame</td>
<td>Beadle</td>
<td>s.n.</td>
<td>26</td>
<td>6</td>
<td>-107</td>
<td>0</td>
</tr>
<tr>
<td>IS8</td>
<td><em>mays</em></td>
<td><em>parviglumis</em></td>
<td>Mexico</td>
<td>Guerrero</td>
<td>Sites 9-10, Teloloapan-Arcelia Hwy</td>
<td>Iltis &amp; Cochrane</td>
<td>81</td>
<td>18</td>
<td>21</td>
<td>-100</td>
<td>12</td>
</tr>
<tr>
<td>IS9</td>
<td><em>mays</em></td>
<td><em>mexicana</em></td>
<td>Mexico</td>
<td>Mexico</td>
<td>km 1.8 WSW of Texcoco</td>
<td>H. Iltis</td>
<td>28620</td>
<td>19</td>
<td>30</td>
<td>-98</td>
<td>55</td>
</tr>
</tbody>
</table>
Fig. 1

Chromosome 1L

IS1
IS2
IS3
IS4
IS5
IS6
IS7
IS8
IS9

200 220 240 260 280 Mb

tb1
Fig. 2

Cupules per rank

Staminate Spikelets

Internode length

Tillering
Fig. 3
Fig. 4
Fig. 5

A  Coding Region

B  Control Region

--- 1 change