Genetics: Early Online, published on May 4, 2015 as 10.1534/genetics.115.175752

Section: Genetics of Complex Traits

Evidence that the origin of naked kernels during maize domestication was caused by a single amino acid substitution in *tga1*

Huai Wang[*,1], Anthony J. Studer[*,2], Qiong Zhao[*,3], Robert Meeley[§] and John F. Doebley[*,4]

Author Affiliations

[*] Laboratory of Genetics, University of Wisconsin, 425 Henry Mall, Madison, WI 53076 U.S.A.

[§] DuPont Pioneer, 7300 NW 62nd Ave., Johnston, IA 50131 U.S.A.

[1] Present address: Monsanto Company, 800 N Lindbergh Blvd., St. Louis, MO 63141 U.S.A.

[2] Present address: Donald Danforth Plant Science Center, 975 N Warson Rd., St. Louis, MO 63132 U.S.A.

[3] Present address: Otsuka America Pharmaceutical, Inc., 2440 Research Blvd., Rockville, MD 20850 U.S.A.

[4] To whom correspondence should be addressed: jdoebley@wisc.edu

Short Title

Amino acid change drives domestication

Key Words

tga1, glume architecture, teosinte, maize, domestication

Corresponding Author

John F. Doebley

Laboratory of Genetics

University of Wisconsin

425 Henry Mall

Madison, WI 53706 USA

tel: 608-265-5803

fax: 608-262-2976

email: jdoebley@wisc.edu

Abstract

teosinte glume architecture1 (tga1), a member of the SBP-box gene family of transcriptional regulators, has been identified as the gene conferring naked kernels in maize vs. encased kernels in its wild progenitor, teosinte. However, the identity of the causative polymorphism within *tga1* that produces these different phenotypes has remained unknown. Using nucleotide diversity data, we show that there is a single fixed nucleotide difference between maize and teosinte in *tga1*, and this difference confers a Lys (teosinte allele) to Asp (maize allele) substitution. This substitution transforms TGA1 into a transcriptional repressor. While both alleles of TGA1 can bind a GTAC motif, maize-TGA1 forms more stable dimers than teosinte-TGA1. Since it is the only fixed difference between maize and teosinte, this alteration in protein function likely underlies the differences in maize and teosinte glume architecture. We previously reported a difference in TGA1 protein abundance between maize and teosinte based on relative signal intensity of a western blot. Here, we show that this signal difference is not due to *tga1* but to a second gene, *neighbor of tga1 (not1)*. *not1* encodes a protein that has 92% amino acid similarity to TGA1 and that is recognized by the TGA1 antibody. Genetic mapping and phenotypic data show that *tga1*, without a contribution from *not1*, controls the difference in covered vs. naked kernels. No trait differences could be associated with the maize vs. teosinte alleles of *not1*. Our results document how morphological evolution can be driven by a simple nucleotide change that alters protein function.

Introduction

Although the study of adaptive evolution through natural or artificial selection dates back to Darwin, the genetic mechanisms that drive changes in morphology remain strongly debated (Stern and Orgogozo 2008). Questions surrounding the number of genes (few or many) and types of mutations (regulatory or coding) in these genes have been a focus in this debate (Carroll 2005, 2008; Hoekstra and Coyne 2007). These questions can only be answered through the genetic and molecular dissection of genes that have undergone selective pressure between lineages or within populations. Crop species offer a powerful system for investigating these questions since crops are the products of continuous directional selection to adapt them to the human controlled environment and human needs (Meyer and Purugganan 2013). Research on crop models is further facilitated by the extensive genetic and genomic resources available for them. Moreover, because of the recent divergence of crop species from their wild progenitors, crop-progenitor pairs remain cross compatible and amenable to genetic analysis.

Maize was domesticated in the central Balsas valley of Mexico ~9,000 years ago from a wild relative called teosinte (Piperno *et al*. 2001; Matsuoka *et al*. 2002). Because the morphology of modern maize is drastically different from teosinte species, the progenitor of maize was highly debated until molecular evidence proved the ancestral link (Mangelsdorf and Reeves 1938; Beadle 1939; Doebley 2001). Remarkably, genes controlling much of the morphological difference between maize and teosinte were shown to map to just six regions of the genome or major domestication loci (Doebley *et al.* 1990). Subsequent studies have sought to elucidate the genetic nature of these loci. Fine mapping of two of these regions, located on opposite arms of chromosome 1, identified causative polymorphisms in the regulatory region of *teosinte branched1* and *grassy tillers1* (Studer *et al.* 2011; Wills *et al.* 2013). These regulatory changes

lead to altered plant architecture, and population genetic data indicate that the regulatory regions of these two genes were targets of selection during domestication. Recently, the domestication locus on chromosome 5 was shown to contain multiple factors (genes) rather than a single large effect gene (Lemmon and Doebley 2014).

Another major maize domestication locus, which is located on chromosome 4, controls whether the grains are enclosed in a "fruitcase" as in teosinte or uncovered as in maize (Dorweiler *et al.* 1993). The fruitcase that encapsulates the teosinte grain is formed from a hardened cup-shaped stem segment (cupule) in which the grain is located and a hardened bract (glume) that seals the grain in the cupule. In maize, the grains are borne naked on the exterior of the ear, and the organs that form the fruitcase in teosinte are redeployed to form the internal central axis of the ear (the cob). The transition from encased to exposed grain greatly facilitated the use of the grain as food. The locus that largely controls this difference has been resolved to a single gene called *teosinte glume architecture1* (*tga1*) (Wang *et al.* 2005). The maize allele of *tga1* disrupts the normal development of the cupulate fruitcase, exposing the grain on the surface of the ear. *tga1* encodes a squamosa-promoter binding protein (SBP), a transcription factor family that has been shown to regulate floral development (Klein *et al.* 1996; Wang *et al.* 2005). Although *tga1* has been identified as the major gene controlling changes in fruitcase development during domestication, the causative polymorphism in *tga1* and how this polymorphism affects the phenotype has not been resolved.

In this manuscript, we show that a single fixed nucleotide polymorphism in the coding sequence of *tga1* distinguishes the maize and teosinte alleles. This difference creates an amino acid substitution that changes TGA1 protein dimerization and alters how TGA1 regulates its targets, with the maize allele acting more as a repressor relative to the teosinte allele. We also

describe the pleiotropic effects of RNAi lines for *tga1* on multiple traits, indicating that *tga1* may play a broad role in development, having effects on kernel size and shape as well as plant architecture. Finally, we describe another gene, *neighbor of tga1* (*not1*), that is tightly liked to *tga1* and has a high sequence homology with *tga1*, but for which we observed no differences in phenotypic effect between the maize and teosinte alleles. Our molecular and genetic analyses of *tga1* show how a simple amino acid change can alter protein function and thereby drive the evolution of a new phenotypic state.

Materials and Methods

Detailed Materials and Methods can be found in Supporting Information.

Plant materials. Maize inbred W22 and W22:*tga1*, an introgression line that contains a teosinte chromosomal segment surrounding *tga1* in a W22 background, were used in most experiments (Dorweiler and Doebley 1997). A set of recombinant lines (T249, T1214, T1464, and T2956) derived from a W22 x W22:*tga1* F_2 fine mapping population that was previously described (Wang *et al.* 2005), were also utilized. The W22:*tga1-ems* line, previously reported in Wang *et al*. (2005), contains the amino acid substitution Leu5 to Asn5, which is immediately upstream of the single amino acid difference between maize- and teosinte-TGA1. The *tga1-ems* allele was recovered from an ethyl methanesulfonate mutagenesis and displays ear phenotypes similar to maize lines containing the teosinte allele of *tga1*. Additional plant materials include the *not1-Mu2* stock from the TUSC collection and the transgenic *tga1-RNAi* lines made at the Plant Transformation Facility at Iowa State University.

Protoplast Transient Assays. Dual luciferase reporter assays were used to determine the repressor function of *tga1* in transient protoplast expression experiments. N-terminal sequences of *tga1* were fused to a GAL4-DNA Binding Domain (DBD) and cotransformed with a firefly luciferase gene downstream of two GAL4 binding sites. Firefly luciferase expression was measured and then normalized to a *Renilla* luciferase internal control. Maize mesophyll protoplasts used for the transient expression experiments were extracted and transformed using a protocol developed by the Sheen Lab (see Supporting Information).

Protein Purification from Plant Tissue. Protein was extracted from ear tissue using a plant total protein extraction kit (Sigma, St. Louis, MO) following the manufacture's instructions. To show that TGA1 forms a dimer *in vivo,* formaldehyde was used to fix immature ear tissue prior to protein extraction.

Binding Assays. Electrophoretic Mobility Shift Assays (EMSAs) for PCR-assisted binding site selection experiments were performed as described previously (Tang and Perry 2003), and for experiments testing the *in vitro* binding of maize-TGA1 to the *not1* promoter, performed as described previously with some modifications (Wang *et al.* 2004). Chromatin immunoprecipitation (ChIP) experiments using an anti-TGA1 polyclonal antibody, were used in to verify the *tga1* bound the *not1* promoter *in vivo*. ChIP assays were performed as described previously (Gendrel *et al.* 2002; Wang *et al.* 2002).

Results

Only one fixed nucleotide difference exists between maize and teosinte alleles of *tga1*. Wang *et al*. (2005) demonstrated by fine-mapping that the causal polymorphism that distinguishes maize and teosinte lies within 1042 bp portion of *tga1* (Genbank: AY883436-AY883460), which includes the first 18 bp of the ORF and 1024 bp upstream of the ATG. With a small sample of maize and teosinte alleles, these authors observed seven nucleotide differences between maize and teosinte: six that are upstream of the start codon, and one that is at position 18 of the ORF. To determine if these seven candidate polymorphisms could be narrowed to a smaller number, we assayed a larger sample (20) of teosinte alleles (Figure S1, Genbank: KR261098- KR261108). These data show that the six upstream polymorphisms no longer represent fixed differences between maize and teosinte but rather that some teosintes possess the same nucleotide as maize at each of these six sites. However, a nucleotide difference at position 18 of the ORF (C for maize and G for teosinte), which encodes a Lys6 to Asp6 substitution from teosinte-TGA1 to maize-TGA1, still remains a fixed difference. Thus, this is the only fixed difference in the causative region that defines the glume architecture difference between maize and teosinte.

***neighbor of tga1* is a tightly linked paralog of *tga1*.** Examination of the genomic region near *tga1* revealed a gene (AC233751.1_FG002) that shares high nucleotide similarity to *tga1*. This gene is located only ~270 kb away from *tga1* and thus it was named *neighbor of tga1 (not1)* (Preston *et al.* 2012). Comparing the TGA1 and NOT1 proteins from maize inbred W22, they have 92% identity in sequence. *not1* also exists in teosinte and a sequence alignment of different alleles of *tga1* and *not1* is shown in Figure S2.

To investigate whether *not1* contributes to the glume architecture difference between teosinte and maize, we investigated the glume phenotype of recombinant inbred lines (RILs) carrying different combinations of the *tga1* and *not1* alleles. RILs that carrying *tga1-teosinte* alleles, such as T249 and W22:*tga1*, all show teosinte glume architecture regardless of whether they have the maize or teosinte allele at *not1* (Table 1). Similarly, RILs that possess the *tga1-maize* allele, such as T1214, T1464, T2956, all show maize glume architecture regardless of whether they have the maize or teosinte allele at *not1*. Thus, the teosinte glume architecture phenotype is associated with *tga1-teosinte* allele but not related to the *not1-teosinte* allele.

We investigated the *tga1* and *not1* gene expression across different lines with RT-qPCR. As shown in Figure 1A, there is no statistical difference in message accumulation between the maize and teosinte alleles of *tga1* (ANOVA: P =0.9467), and *tga1* message level is not affected by the genotype at *not1*. However, *not1* message level is lowest when there are maize alleles at both *not1* and *tga1*, highest when there are teosinte alleles at both genes, and intermediate for the heteroallelic genotypes. These results indicate that the maize allele of *not1* accumulates less message than the teosinte allele, and they suggest that maize-TGA1 represses *not1* expression compare to teosinte-TGA1.

An antibody was generated using the TGA1 C-terminal protein and was used for a western blot as part of the initial characterization of *tga1* (Wang *et al.* 2005). On the western blot, the signal associated with protein samples for genetic stocks with the *tga1-teosinte* allele were stronger than for stocks with the *tga1-maize* allele (Wang *et al.* 2005). However, these stocks not only differ for their *tga1* allele but they also differ for their alleles at *not1*. This situation raises the possibility that the difference in signal strength on the western blot was due to *not1* rather

than *tga1*. If the Anti-TGA1 antibody cross reacts with NOT1, then the western signal might be a combination of both TGA1 and NOT1.

To investigate this possibility, we performed a western blot with proteins from comparable developmentally staged ears of our different RILs. The strongest signals were detected from the ears of all the RILs that carry a *not1-teosinte* allele including W22:*tga1*, T1214, T1464 and T2956 (Figure 1B). By comparison, line (T249) that contain a teosinte allele at *tga1* but a maize allele at *not1*, has a signal intensity on the western blot that is much less. In addition, the western blot signals from W22:*tga1-ems* that carries a *tga1-ems* allele and a *not1-maize allele* do not show dramatic difference in signal from W22. These observations suggest that strong signals in western blots are associated with *not1-teosinte* instead of *tga1-teosinte*. Thus, the observation made by Wang *et al.* (2005) that *tga1-teosinte* allele confers greater protein accumulation than the *tga1-maize* allele does not appear to be correct. Rather, the level of protein accumulation associated with these two alleles appears roughly equivalent.

We further investigated the TGA1 and NOT1 proteins by performing gel electrophoresis for an extended period of time to see if the TGA1 and NOT1 proteins could be separated (Figure 1C). In this analysis, we included an additional *not1* allele with a *Mu* element insertion (*not1-Mu2*) for which RT-qPCR shows no evidence of a transcript (Figure S3). The RILs with *not1-maize* all show two protein bands correspond to TGA1 and NOT1. The *not1-Mu2* line is missing the lower of these two bands. Thus, the lower band appears to be NOT1. The RILs with *not1-teosinte* show a single thick band which could be co-migrating TGA1 and NOT1 proteins. Importantly, a comparison of RIL T249 and W22 shows that the band corresponding to NOT1 is stronger when *tga1-teosinte* is present than when *tga1-maize* is present. This observation is consistent with our interpretation of Figure 1B that the stronger protein signal for the W22:*tga1*

line represents NOT1 rather than TGA1. The combined RT-qPCR and western data suggest that *not1-teosinte* is expressed higher than *not1-maize*, and that *tga1-maize* represses *not1* such that the greatest difference in *not1* expression is seen between lines that are *tga1-maize; not1-maize* versus *tga1-teosinte; not1-teosinte* (Figure 1A, C, B). Again, the statement by Wang *et al.* (2005) that *tga1-teosinte* allele confers greater protein accumulation than the *tga1-maize* allele does not appear to be correct.

Lys6 to Asp6 substitution converted TGA1 N-terminal into a repression domain. Based on the maize-teosinte sequence comparison and lack of an expression difference between the maize and teosinte alleles of *tga1*, our working hypothesis is that the single amino acid substitution (Lys to Asp) at the 6th position of *tga1* is the causative site for the lost of teosinte glume architecture during maize domestication. To test if the Lys6 to Asp6 substitution controls the functional difference between maize- and teosinte-TGA1, we employed a protoplast transient assay system**.**

We constructed six effectors and two reporter constructs for the transient assays (Figure 2A, B). We mixed plasmids with different effector and reporter combinations (Figure 2C) and introduced the mixture into maize mesophyll protoplasts by electroporation. After incubating the transfected protoplasts for 16-18 hrs, the firefly luciferase activity was normalized to the *Renilla* luciferase activity for each assay. Transfection of the LD-VP16 effector along with the reporters gave strong firefly gene expression (Figure 2C); however, all other effectors (maize-GD, teosinte-GD, ems-GD) along with the reporters in absence of LD-VP16 showed no activity in reporter gene expression relative to the negative control effector (GD) (Figure 2C, light gray

columns). These results suggest TGA1 N-terminal domain is not an activation domain for any of the three alleles.

Co-transfection of LD-VP16 with an effector encoding the GAL4-DBD fused to the IAA17(I) (positive control for repression) resulted in repressed firefly expression (Figure 2C). Co-transfection of LD-VP16 with the effector carrying the maize allele TGA1 N terminal fusion also showed repressed firefly luciferase activity. Interestingly, co-transfection with the TGA1 N terminal domain from teosinte or ems allele effectors showed similar reporter activities to the GD effector (negative control). These results indicate that the N terminal domain of maize-TGA1 is an active repression domain, but the N-terminal of teosinte-TGA1 or ems-TGA1 are not active repressors. The results also suggest that the repression activity of TGA1 N-terminal domain is established by just a single naturally occurring amino acid substitution (Lys6 to Asp6) and it was reversed by the ems induced substitution (Leu5 to Phe5).

PCR-assisted Binding site selection. To assay whether the Lys6 to Asp6 substitution affects the DNA binding activity of the protein and to better understanding TGA1 function, we performed PCR-assisted binding site selection to determine the binding site of TGA1. Using Electrophoretic Mobility Shift Assays (EMSAs), we observed that both maize- and teosinte-TGA1 can shift up two bands but their patterns are opposite (Figure 3). Maize-TGA1 shifted up a stronger upper band and teosinte-TGA1 has a stronger lower band. However, sequence consensus of DNA fragments derived from these shifted bands showed that they all contain a GTAC motif. This is consistent with previous reports that SBP-domain proteins bind to GTAC (Birkenbihl *et al.* 2005; Kropat *et al.* 2005). These results indicate that the Lys6 to Asp6 substitution did not affect the

binding site specificity of TGA1 for the GTAC motif, but likely affected the configuration of TGA1 binding to DNA.

TGA1 binds to the promoter of *not1 in vitro* and *in vivo*. We isolated the promoter segments of *not1-maize* and *not1-teosinte* alleles (Figure S4, Genbank: KR261109 and KR261110). Interestingly, both sequences share a conservative region that contains the GTAC motif. We took partial sequence surrounding this GTAC motif to make a probe for EMSA (Figure 4A). Our results show this sequence from the *not1* promoter can bind maize-TGA1 *in vitro* (Figure 4B). Furthermore, the binding activity of maize-TGA1 to this DNA fragment was abolished totally with the GTAC motif mutated to CTAC (Figure 4B). These results suggest that TGA1 binds to the *not1* promoter at the GTAC motif.

A Chromatin immunoprecipitation (ChIP) assay using an anti-TGA1antibody was performed to confirm the *in vivo* interaction between TGA1 and *not1* promoter. Because anti-TGA1 can also cross action with NOT1, we used tissue from plants with the *not1-Mu2* allele, which is a null mutant for *not1*. The ChIP-PCR results showed that the DNA fragment of the *not1* promoter region was enriched in the anti-TGA1 ChIP population as compared to the input control (Figure 4C). Furthermore, ChIP populations generated using non- or pre-immune serum did not show specific selection of the *not1* fragment. To quantify the degree of enrichment, we also performed qPCR which shows the *not1* promoter region was enriched about 4.8-fold in the anti-TGA1 population (ChIP=4.68 ± 0.27; input control=0.97 ± 0.11) (Figure 4D).

Maize-TGA1 forms more stable dimers than teosinte-TGA1. The DNA probes used for EMSA in Figures 3 and 4 are different, but they both showed doublet bands. Thus, the double

banding pattern is not likely related to the DNA probe sequences. To determine if the doublet was caused by protein dimerization, we performed additional EMSA using tagged and tag-free versions of TGA1 that are expect to migrate differentially during electrophoresis and thus be distinguishable. In addition to these two full-length versions of TGA1, we also used a shortened version of TGA1 with the N-terminus removed.

As shown in Figure 5A, all the full length TGA1 proteins, including tag-maize-TGA1, tag-teosinte-TGA1, tag-ems-TGA1 and maize-TGA1, shifted up two bands. The upper band is the major band when using tag-maize-TGA1 or maize-TGA1 while the lower band is the major band when using tag-teosinte-TGA1 or tag-ems-TGA1. These results suggest that TGA1 exists as dimers and monomers dynamically, and that maize-TGA1 tends to form more stable dimers.

In Figure 5A, one can also see that the DNA-protein complexes with tag-maize-TGA1 (version 1) run slower in the gel than the complex from tag-free maize-TGA1 (version 2). When we mixed tag-maize-TGA1 and tag-free maize-TGA1 together in the same binding reactions, we found that a novel intermediated band was detected. We interpret this new band as a heterodimer of tag-maize-TGA1/tag-free maize-TGA1 proteins. Intriguingly, a truncated TGA1 with 103 amino acids removed from N-terminus only shifts up a single band. These results indicate that TGA1 can form dimers and that the N-terminus of TGA1 is necessary for dimerization. We conclude that the Lys6 to Asp6 mutation, which is located in the TGA1 N-terminus, can affect the dynamic ratio between dimers and monomers.

To investigate whether TGA1 forms dimers *in vivo*, we performed a western blot assay using formaldehyde cross-linked ear tissue from *not1-Mu2* plants. Formaldehyde can covalently preserve protein dimers *in vivo*, thus prevent the dimer disassociation during protein preparation and electrophoresis (Sang *et al.* 2005). As shown in Figure 5B, only a single band was detected

at ~50 kD from non-fixed tissue, and this band is approximately the size of TGA1 as monomer. However, when using cross-linked tissue, an additional band which is approximately twice the weight of the TGA1 monomer was recognized by the anti-TGA1 antibody. These results suggest that maize-TGA1 forms homodimers *in vivo,* which is consistent with our *in vitro* assay data.

***tga1/not1* loss-of-function plants via RNAi and their phenotype.** To understand the function of *tga1*, we screened the available maize mutant collections for *tga1* loss-of-function alleles without success (see materials and methods). We did not recover any mutant alleles for *tga1*, however we found 5 *Mu* insertions in *not1*. The *not1-Mu2* allele is a null mutation lacking detectable expression (Figure S3), although we did not observe any phenotypic difference between homozygous *not1-Mu2* plants and wildtype sibs in a segregating population.

We attempted to generate transgenic *tga1* overexpression plants by putting *tga1-maize* or *tga1-teosinte* under a constitutive rice *Actin1* promoter. Surprisingly, no *tga1* overexpression transgenic plants could be generated with either the maize or teosinte allele. *tga1* contains a miRNA156 target site, and thus miRNA156 could inhibit the overexpression. To rule out this possibility, we built two new constructs with the miRNA156 site in *tga1* mutated synonymously and placed them under a maize *Ubiqitin* promoter. However, we still could not produce transgenic plants and there was difficulty even obtaining transgenic callus with these vectors. Other control constructs that were transformed side-by-side worked fine as reported to us by Dr. Kan Wang at the Plant Transformation Facility, Iowa state University. These results suggest that the overexpression of *tga1* inhibits some aspect of plant development and thus plant regeneration during the transformation process.

We next used an RNAi based approach to generate *tga1* loss-of-function plants. Because RNAi is achieved by overexpression of a hairpin structure (Kusaba 2004), it is likely not efficient enough to generate a complete knock out *tga1* mRNA entirely, especially during plant regeneration. Using this approach we were able to recover *tga1-RNAi* transgenic plants. Five of the *tga1-RNAi* events were grown to maturity and crossed to maize inbred W22. Then, 60 of progeny plants from each cross were grown and the segregation of transgene identified by a BASTA leaf painting assay. Progeny groups for four events showed an approximate 1:1 ratio for BASTA resistance: susceptible plants, suggesting that these 4 events have single-insertions.

As shown in Figure 6, *tga1-RNAi* plants present some interesting phenotypes. The *tga1/not1* loss-of-function plants have much longer lateral branches (Figure 6A, B, C). A few lateral branches even have secondary ears developed along them (Figure 6C). The glumes of *tga1-RNAi* plants are enlarged, which is a characterization of W22:*tga1* glumes (Figure 6F, G, J, K). However, these glumes are relatively paperish and less thick, less hard and less polished. The kernels from the *tga1-RNAi* plants are narrower (Figure 6I) comparing to control kernels (Figure 6H). Furthermore, they have a pointed tip which makes the ear “prickly” (Figure 6M), while the control ear is smooth (Figure 6L). Interestingly, the *tga1-RNAi* plants have prop roots developed at the first 4-6 nodes of the stalk, while the control plants only have two nodes with prop roots (Figure 6D and E). This collection of phenotypes suggests that *tga1/not1* not only have function in glume architecture, but also function in juvenile growth, lateral branch formation, and ear development. The fact that the *tga1-RNAi* plants display some characteristics of the *tga1-teosinte* glume phenotype suggests that the Lys6 to Asp6 from teosinte to maize is a gain-of-function mutation.

Quantitative traits associated with *tga1-RNAi*. Statistically significant associations were identified between the *tga1-RNAi* transgene and some quantitative traits related to plant and ear architecture. We compared plants with and without the *tga1-RNAi* transgene as determined by a BASTA panting assay. We analyzed 30 resistance (TG=transgene present) and 30 susceptible (NTG=transgene absent) plants each from segregating F_2 families for four events. Associations between the phenotypes and *tga1-RNAi* transgene genotype were tested using T-tests.

As shown in Figure 7, the *tga1-RNAi* transgene is associated with statistically significant effects on lateral branch number, length of the uppermost lateral branch, the length of the blade of the first husk leaf, and the number of nodes with prop roots. The results suggest that TGA1/NOT1 represses the growth of lateral branches in length and numbers. The association of some ear traits with *tga1-RNAi* was also observed (Figure 7). Glume length was significantly increased when TGA1/NOT1 were knocked down/out. In contrast, ear diameter and the weight of 50 kernels decreased in *tga1-RNAi* plants. We also measured some additional ear traits, such as ear length and kernel row number, however, there were no statistically significant associations between these traits and *tga1-RNAi* transgene.

In summary, these results suggest that TGA1/NOT1 have broader functions beyond controlling glume architecture in maize and teosinte. Specifically, they may also affect traits such as seed shape, seed weight, ear and lateral branch morphology as well as the juvenile seedling phase of plant growth.

Discussion

A previous population genetics analysis of the *tga1* identified seven fixed differences using a small sample of maize and teosinte accessions (Wang *et al.* 2005). Here, we used a larger sample

size of maize and teosinte, which revealed that six of these seven nucleotide sites are polymorphic in both maize and teosinte (Figure S1). With this larger sample, only a single fixed polymorphism was identified between maize and teosinte within *tga1*. This single site encodes and Lys6 to Asp6 substitution between teosinte and maize at the sixth amino acid position from the N-terminus of the protein. This elevates this Lys6 to Asp6 substitution as the most likely candidate for the causative difference between maize and teosinte in *tga1*.

We tested the functional consequences of the Lys6 to Asp6 substitution using protoplast transient assays (Tiwari *et al.* 2004). These assays revealed that this single amino acid substitution alters the function of the TGA1 protein. The Lys6 to Asp6 substitution gives the maize-TGA1 a strong repressor function as seen by reporter expression levels that are comparable to the repressor control (Figure 2). Furthermore, this amino acid substitution also increases dimerization of TGA1, as shown by gel shift assays (Figure 5). The increased dimer formation and repressor function of the maize-TGA1 are likely concomitants. Arguably, the most well characterized repressor, the *lac* repressor (LacR) in *E. coli*, forms a stable dimer of dimers, which then produces a functional tetramer through weak associations (reviewed in Lewis 2005). Amino acid substitutions that weaken dimer formation of LacR eliminate repressor function. Furthermore, a single amino acid substitution has been shown to block dimerization (Dong *et al.* 1999; Spott and Dong 2000). Thus, the observed correlation of dimerization and repressor strength reported here is consistent with functional studies in other systems.

The binding site of TGA1 was determined using EMSA and shown to be GTAC (Figure 3). Both the maize- and teosinte-TGA1 were used, however, no difference in binding site preference was observed. Thus, the amino acid substitution that leads to repressor function is likely the mechanism by which plant morphology is altered, not differences in binding site

specificity. The *tga1* binding motif was found to bind the promoter of *not1* both *in vitro* and *in vivo*, suggesting that the *not1* and *tga1* may function in the same pathways. This result begins to unravel a potential cascade of gene expression changes that accompanies the alteration of a major domestication gene.

tga1 was discovered based on its specific effects of the maize vs. teosinte allele on the development of the teosinte fruitcase and maize cob (Dorweiler *et al.* 1993). Although these alleles differ in functional amino acid substitution, both alleles are expressed at comparable levels (Wang *et al.* 2005). We assayed the broader effects of *tga1* by using an RNAi construct to reduce or eliminated *tga1* gene expression. Maize lines expressing an RNAi construct targeting *tga1* displayed pleiotropic morphological effects on several branching and kernel traits (Figures 6, 7), which had not previously been associated with this domestication gene (most recently Brown *et al.* 2011). With regard to branching, these RNAi lines likely remove the repressive function of TGA1/NOT1, allowing the outgrowth of axillary branches. The effects on kernel shape and size may be related to the fact that the kernel resides within the fruitcase in teosinte, and thus, fruitcase and kernel development are coordinately regulated by *tga1*. Whether the maize vs. teosinte allele of *tga1* affect any of these additional traits is unknown, but such effects have not previously been reported (Dorweiler and Doebley 1997).

An open question is whether the effects of the RNAi construct on kernel and branching traits results from a knock-down of *tga1* or *not1*. The RNAi construct was generated using *tga1* sequence, but given the sequence similarity between *tga1* and *not1*, it is likely that the RNAi construct targets both genes. However, given that the *Mu* insertions in *not1* did not produce the morphological phenotypes seen in the RNAi lines, we infer that the phenotypes observed in the RNAi lines are either attributable to *tga1*, or a redundant function of *tga1* and *not1* and thus, can

only be observed when both are knocked down. However, without a null *tga1* allele, we cannot show concussively that these phenotypes are specific to *tga1*. These inferences about *tga1* vs. *not1* function are further complicated by the binding of TGA1 to the *not1* promoter. While multiple morphological phenotypes are observed in the RNAi lines, the single amino acid substitution fixed during domestication seems to be specific to the ear traits reported previously (Wang *et al.* 2005).

Both *tga1* and *not1* belong to the squamosa-promoter binding protein (SBP) family of transcription factors. Members of this family have been shown to regulate meristem development, and manipulations of these regulators have produced both plant architecture and ear phenotypes (Chuck *et al.* 2010, 2014). While there is a clear homolog for *tga1* in other grass species, the *tga1/not1* duplication occurred at the base of the *Zea* genus, evident by the absence of this duplication in other lineage (Preston *et al*. 2012). The presence of the *tga1/not1* duplication in maize may have facilitated the subfunctionalization of *tga1/not1* such that *tga1* alone controls the fruitcase/cob in teosinte/maize while *tga1* functions in a redundant manner with *not1* to regulate plant architecture traits. This would explain the phenotypes observed in the RNAi lines, which likely target both *tga1*/*not1*, when no phenotype was present in *not1* mutant plants. Furthermore, various *tga1* alleles have only been reported to display ear phenotypes (Wang *et al.* 2005; Brown *et al.* 2011). This hypothesis is further supported by work on the ortholog of *tga1* in rice (LOC_Os08g41940; *OsSPL16*), which has pleiotropic plant and ear phenotypes (Wang *et al.* 2011, 2012).

In this paper, we have shown that an amino acid substitution in *tga1* is the causal variant that underlies the origin of the naked grains of maize as compared to the covered grains of teosinte. Although the predominant mechanism for morphological evolution may be alterations

in gene expression (Carroll 2008), changes in protein function are also involved as shown here. Investigation of how protein evolution contributes to the evolution of new morphological forms enhance our understanding of how adaptions arise.

Acknowledgement

We thank Bao Kim Nugyen, Jesse Rucker, Tina Nussbaum Wagler and Lisa Kursel for technical assistance. This work supported by National Science Foundation grants IOS1025869 and IOS1238014, and a USDA-Hatch grant MSN169062.

Literature Cited

Beadle G. W., 1939 Teosinte and the origin of maize. J. Hered. 30: 245–247.

Birkenbihl R. P., G. Jach, H. Saedler, and P. Huijser, 2005 Functional dissection of the plant-specific SBP-domain: overlap of the DNA-binding and nuclear localization domains. J. Mol. Biol. 352: 585–596.

Brown P. J., N. Upadyayula, G. S. Mahone, F. Tian, P. J. Bradbury et al., 2011 Distinct genetic architectures for male and female inflorescence traits of maize. PLoS Genet. 7: e1002383.

Carroll S. B., 2005 Evolution at two levels: On genes and form. PLoS Biol. 3: 1159–1166.

Carroll S. B., 2008 Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. Cell 134: 25–36.

Chuck G., C. Whipple, D. Jackson, and S. Hake., 2010 The maize SBP-box transcription factor encoded by tasselsheath4 regulates bract development and the establishment of meristem boundaries. Development 137: 1243–1250.

Chuck G. S., P. J. Brown, R. Meeley, and S. Hake, 2014 Maize SBP-box transcription factors unbranched2 and unbranched3 affect yield traits by regulating the rate of lateral primordia initiation. Proc. Natl. Acad. Sci. USA. 111: 18775–18780.

Doebley J., A. Stec, J. Wendelt, and M. Edwardst, 1990 Genetic and morphological analysis of a maize-teosinte F2 population: Implications for the origin of maize. Proc. Natl. Acad. Sci. USA. 87: 9888–9892.

Doebley J., 2001 George Beadle's other hypothesis: one-gene, one-trait. Genetics 158: 487–493.

Dong F., S. Spott, O. Zimmermann, B. Kisters-woike, B. Mu, et al., 1999 Dimerisation Mutants of Lac Repressor . I . A monomeric Mutant, L251A, that Binds lac Operator DNA as a Dimer. J. Mol. Biol. 290: 653–666.

Dorweiler J., A. Stec, J. Kermicle, and J. Doebley, 1993 Teosinte glume architecture 1: A Genetic Locus Controlling a Key Step in Maize Evolution. Science 262: 233–235.

Dorweiler J., and J. Doebley, 1997 Developmental analysis of teosinte glume architecture1: A key locus in the evolution of maize (Poaceae). Am. J. Bot. 84: 1313–1322.

Gendrel A., Z. Lippman, and C. Yordan, 2002 Dependence of heterochromatic histone H3 methylation patterns on the Arabidopsis gene DDM1. Science. 297: 1871–1873.

Hoekstra H. E., and J. Coyne, 2007 The locus of evolution: evo devo and the genetics of adaptation. Evolution 61: 995–1016.

Klein J., H. Saedler, and P. Huijser, 1996 A new family of DNA binding proteins includes putative transcriptional regulators of the Antirrhinum majus floral meristem identity gene SQUAMOSA. Mol. Gen. Genet. 250: 7–16.

Kropat J., S. Tottey, R. P. Birkenbihl, N. Depège, and P. Huijser, et al., 2005 A regulator of nutritional copper signaling in Chlamydomonas is an SBP domain protein that recognizes the GTAC core of copper response element. Proc. Natl. Acad. Sci. USA. 102: 18730–18735.

Kusaba M., 2004 RNA interference in crop plants. Curr. Opin. Biotechnol. 15: 139–143.

Lemmon Z. H., and J. F. Doebley, 2014 Genetic Dissection of a Genomic Region with Pleiotropic Effects on Domestication Traits in Maize Reveals Multiple Linked QTL. Genetics. 198: 345-353.

Lewis M., 2005 The lac repressor. CR. Biol. 328: 521–548.

Mangelsdorf P. C., and R. G. Reeves, 1938 The origin of maize. Proc. Natl. Acad. Sci. USA. 24: 303–312.

Matsuoka Y., Y. Vigouroux, M. M. Goodman, G. J. Sanchez, E. Buckler, et al., 2002 A single domestication for maize shown by multilocus microsatellite genotyping. Proc. Natl. Acad. Sci. USA. 99: 6080–6084.

Meyer R. S., and M. D. Purugganan, 2013 Evolution of crop species: genetics of domestication and diversification. Nat. Rev. Genet. 14: 840–852.

Piperno D. R., K. V. Flannery, and N. Cave, 2001 The earliest archaeological maize (*Zea mays L.*) from highland Mexico: New accelerator mass spectrometry. Proc. Natl. Acad. Sci. USA. 98: 2101–2103.

Preston, J. C., H. Wang, L. Kursel, J. Doebley, and E. A. Kellogg. 2012 The role of teosinte glume architecture (tga1) in coordinated regulation and evolution of grass glumes and inflorescence axes. New Phytol. 193: 204–215.

Sang Y., Q. Li, V. Rubio, Y. Zhang, J. Mao, et al., 2005 N-Terminal Domain – Mediated Homodimerization Is Required for Photoreceptor Activity of Arabidopsis. Plant Cell 17: 1569–1584.

Spott S., and F. Dong, 2000 Dimerisation mutants of Lac repressor. II. A single amino acid substitution, D278L, changes the specificity of dimerisation. J. Mol. Biol. 296: 673–684.

Stern D. L., and V. Orgogozo, 2008 The loci of evolution: how predictable is genetic evolution? Evolution 62: 2155–2177.

Studer A., Q. Zhao, J. Ross-Ibarra, and J. Doebley, 2011 Identification of a functional transposon insertion in the maize domestication gene tb1. Nat. Genet. 43: 1160–1163.

Tang W., and S. E. Perry, 2003 Binding site selection for the plant MADS domain protein AGL15: an in vitro and in vivo study. J. Biol. Chem. 278: 28154–28159.

Tiwari S. B., G. Hagen, and T. J. Guilfoyle, 2004 Aux/IAA Proteins Contain a Potent Transcriptional Repression Domain. Plant Cell. 16: 533–543.

Wang H., W. Tang, C. Zhu, and S. Perry, 2002 A chromatin immunoprecipitation (ChIP) approach to isolate genes regulated by AGL15, a MADS domain protein that preferentially accumulates in embryos. Plant J. 32: 831–843.

Wang H., L. V. Caruso, A. B. Downie, and S. E. Perry, 2004 The Embryo MADS Domain Protein AGAMOUS-Like 15 Directly Regulates Expression of a Gene Encoding an Enzyme Involved in Gibberellin Metabolism. Plant Cell 16: 1206–1219.

Wang H., T. Nussbaum-Wagler, B. Li, Q. Zhao, Y. Vigouroux, et al., 2005 The origin of the naked grains of maize. Nature 436: 714–719.

Wang S. S., C. S. Wang, T. H. Tseng, Y. L. Hou, and K. Y. Chen, 2011 High-resolution genetic mapping and candidate gene identification of the SLP1 locus that controls glume development in rice. Theor. Appl. Genet. 122: 1489–1496.

Wang S., K. Wu, Q. Yuan, X. Liu, Z. Liu, et al., 2012 Control of grain size, shape and quality by OsSPL16 in rice. Nat. Genet. 44: 950–954.

Wills D. M., C. J. Whipple, S. Takuno, L. E. Kursel, L. M. Shannon, et al., 2013 From Many, One: Genetic Control of Prolificacy during Maize Domestication. PLoS Genet. 9: e1003604.

Table 1. Glume architecture phenotypes of isogenic lines with different genotypes at *tga1* and *not1*

Genotype	Allele at *tga1*	Allele at *not1*	Glume architecture trait
W22	maize	maize	maize like
T249	teosinte	maize	teosinte like
T1214	maize	teosinte	maize like
T1464	teosinte	teosinte	teosinte like
T2956	maize	teosinte	maize like
W22: *tga1*	teosinte	teosinte	teosinte like
W22:*tga1-ems*	ems	maize	teosinte

Figure Captions

Figure 1. *tga1* and *not1* gene expression and protein accumulation. (A) RT-qPCR results showing *tga1* (top panel) and *not1* (bottom panel) expression. *tga1* expression was not statistically different between genetic stocks (ANOVA: P =0.9467, n = 10). *not1* expresses less when *tga1* is a maize allele. Maize-*not1* expresses less in W22 than in T249 and EMS. Teosinte-*not1* expresses less in T1464 than in TGA. (B) Anti-TGA1 recognizes both TGA1 and NOT1, however, TGA1 and NOT1 can not be separated with a short gel run performed in duplicate. (C) Teosinte-NOT1 cannot be separated from TGA1, but maize-NOT1 can be resolved under extended gel running conditions. The allelic compliment of each genetic stock is indicated *M*, maize allele; *t*, teosinte-allele; *ems*, ems allele.

Figure 2. Protoplast transient dual-luciferase assays show maize-TGA is a transcriptional repressor. (A) Diagram showing effector construct components. (B) Diagram showing reporter construct components. (C) Effectors 1-5 were co-transformed with both reporters into maize protoplasts with and without the activation control (effector 6). Dual-luciferase assay results are shown in Relative Light Units (RLU), y-axis, (± SE; n = 6). Maize-TGA1 (effector 2) shows stronger repression than the repression control (effector 5).

Figure 3. Binding site selection for maize-TGA1 and teosinte-TGA1. α-^{32}P labeled DNA fragments that contain 26 randomized nucleotides were incubated with maize or teostine-TGA1. After 5 rounds of EMSAs, the shifted DNA bands were excised, and the consensuses binding motif was obtained. Both maize-TGA1 and teosinte-TGA1 proteins bind to DNA fragments with

a GTAC core motif. For each allele 50ng, 100ng, and 200ng of TGA1 protein are shown (left to right).

Figure 4. TGA1 binds the *not1* promoter *in vitro* and *in vivo* through GTAC core motif. (A) The *not1* promoter DNA probe sequence containing the GTAC motif. (B) EMSA showing that maize-TGA1 binds to the *not1* promoter *in vitro* through the GTAC motif. The *not1* probe shifts up two bands with maize-TGA1 protein, but the band shifting is abolished when the GTAC core is mutated to CTAC. (C-D) Chromatin immunoprecipitation (ChIP) confirm TGA1 binds the *not1* promoter *in vivo*. (C) RT-PCR amplification of the *not1* promoter region containing the GTAC core motif. Input, total input chromatin DNA before precipitation; ChIP, chromatin DNA precipitated with anti-TGA1 antibody; PI, DNA precipitated with pre-immune serum; mock, no antibody or serum added. (D) RT-qPCR showing the relative enrichment of *not1* promoter in the anti-TGA1 ChIPs (± SE; $n = 7$).

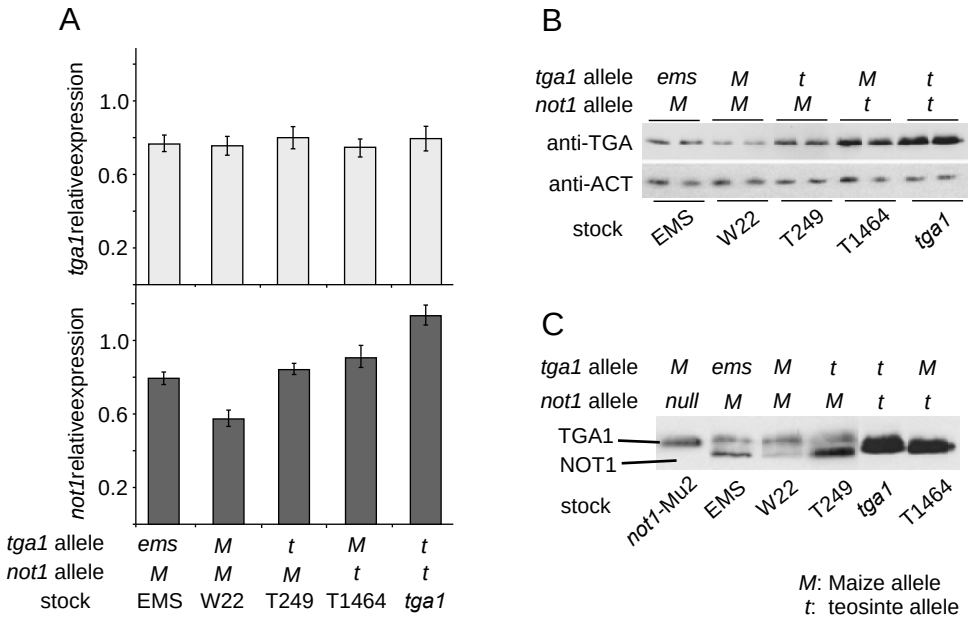
Figure 5. TGA1 forms a dimer *in vitro* and *in vivo*. (A) An EMSA showing that maize-TGA1 forms more stable dimer than teosinte- or ems-TGA1. Removing the N-terminal domain of TGA1 abolished the dimerization of TGA1, shown in lane 2. A mixture of maize-TGA1 proteins with/without tags for EMSA confirmed TGA1 forms homo and heterodimers (lane 6). (B) Protein from formaldehyde cross-linked ears produce two bands on a western blot at approximately 50 kD and 100 kD, while protein from unfixed ears only have the lower band.

Figure 6. Phenotypes of *tga-RNAi* plants. (A-C) *tga1-RNAi* plants have longer lateral branches. (A) shows a non-transgenic control plant with a short lateral branch. (B) shows a typical *tga1-*

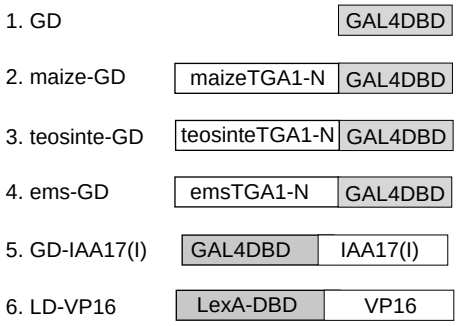
RNAi plant from event 4 with a long lateral branch. (C) shows lateral branches with leaves removed, top one is from non-transgenic control, the others are from *tga1-RNAi* plants which have long shanks and some even have secondary ears. (D-E) a typical *tga1-RNAi* plant (E) has prop roots on more nodes than a control plant (D). (F-M) shows the ear and kernel phenotypes. G, I, K and M are from *tga1-RNAi*. F, H, J and L are controls. Compared to the control, *tga1-RNAi* ears have enlarged glumes (G, K) and narrow shaped seeds (I). Furthermore, each *tga1-RNAi* kernel is pointed (M) on the surface, which results in a spiky ear.

Figure 7. Quantitative effects associated with the *tga1-RNAi* lines. Trait values were plotted on the y-axis and dots represent the value range for each phenotype. Diamonds are centered on the mean values for the traits. Data from four events are presented with TG signifying plants that contain the *tga1-RNAi* transgene and NTG for non-transgenic plants. The *P*-value for each T-test is displayed in the upper right corner of each graph.

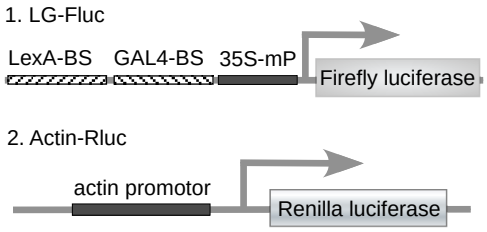
Figure 1


A
tga1 relative expression
1.0
0.6
0.2
not1 relative expression
1.0
0.6
0.2
tga1 allele ems M t M t
not1 allele M M M t t
stock EMS W22 T249 T1464 tga1
B
tga1 allele ems M t M t
not1 allele M M M t t
anti-TGA
anti-ACT
stock EMS W22 T249 T1464 tga1
C
tga1 allele M ems M t t M
not1 allele null M M M t t
TGA1
NOT1
stock not1-Mu2 EMS W22 T249 tga1 T1464
M: Maize allele
t: teosinte allele

A. Effectors



B. Reportors



C. Dual-luciferase assay

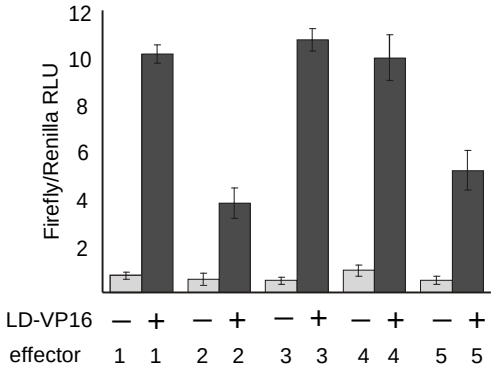


Figure 2

Figure 3

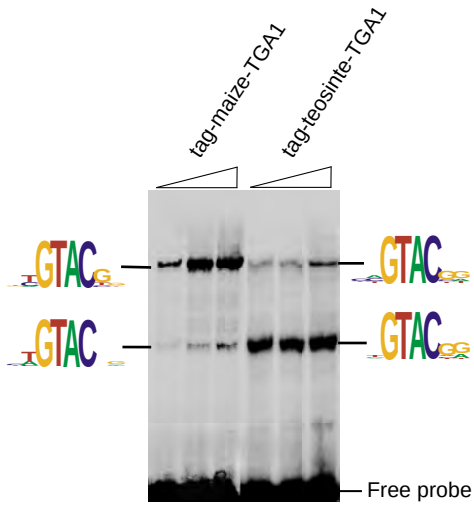

tag-maize-TGA1
tag-teosinte-TGA1
GTAC
GTAC
GTAC
GTAC
Free probe

Figure 4

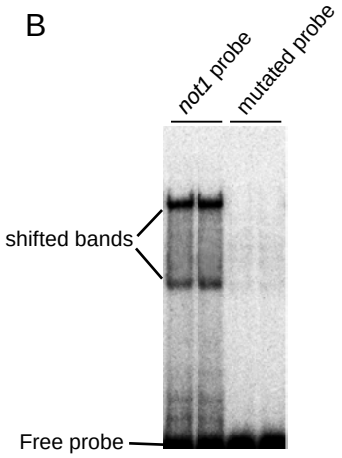
A

Probe from *not1* promotor

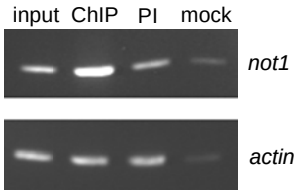
-720 TTGCTACAGTCGCAGCTTCTG
TCTGCAAA***GTAC***TGACTGCTCCC
ACTCCACACCCCCAGTTCCG -654

Mutated probe: GTAC to CTAC

B



C



D

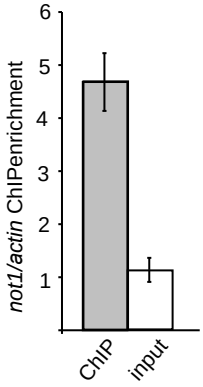


Figure 5

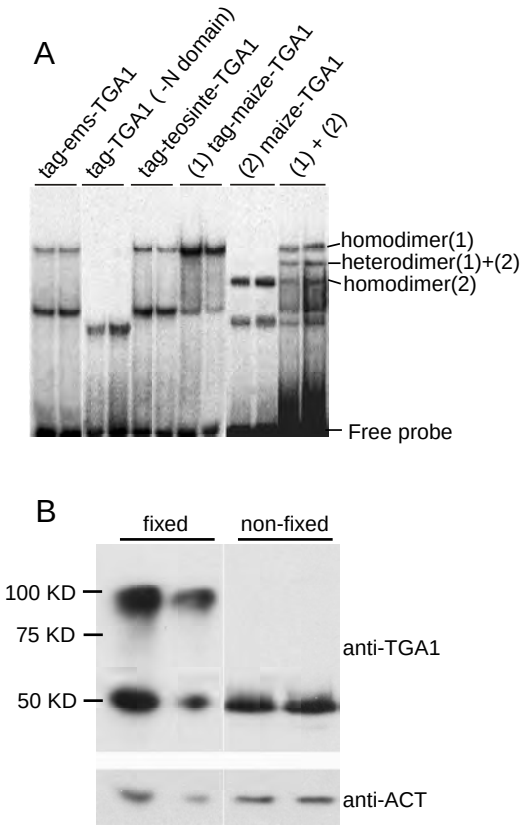

A
tag-ems-TGA1
tag-TGA1 (-N domain)
tag-teosinte-TGA1
(1) tag-maize-TGA1
(2) maize-TGA1
(1) + (2)
homodimer(1)
heterodimer(1)+(2)
homodimer(2)
Free probe
B
fixed
non-fixed
100 KD
75 KD
50 KD
anti-TGA1
anti-ACT

Figure 6

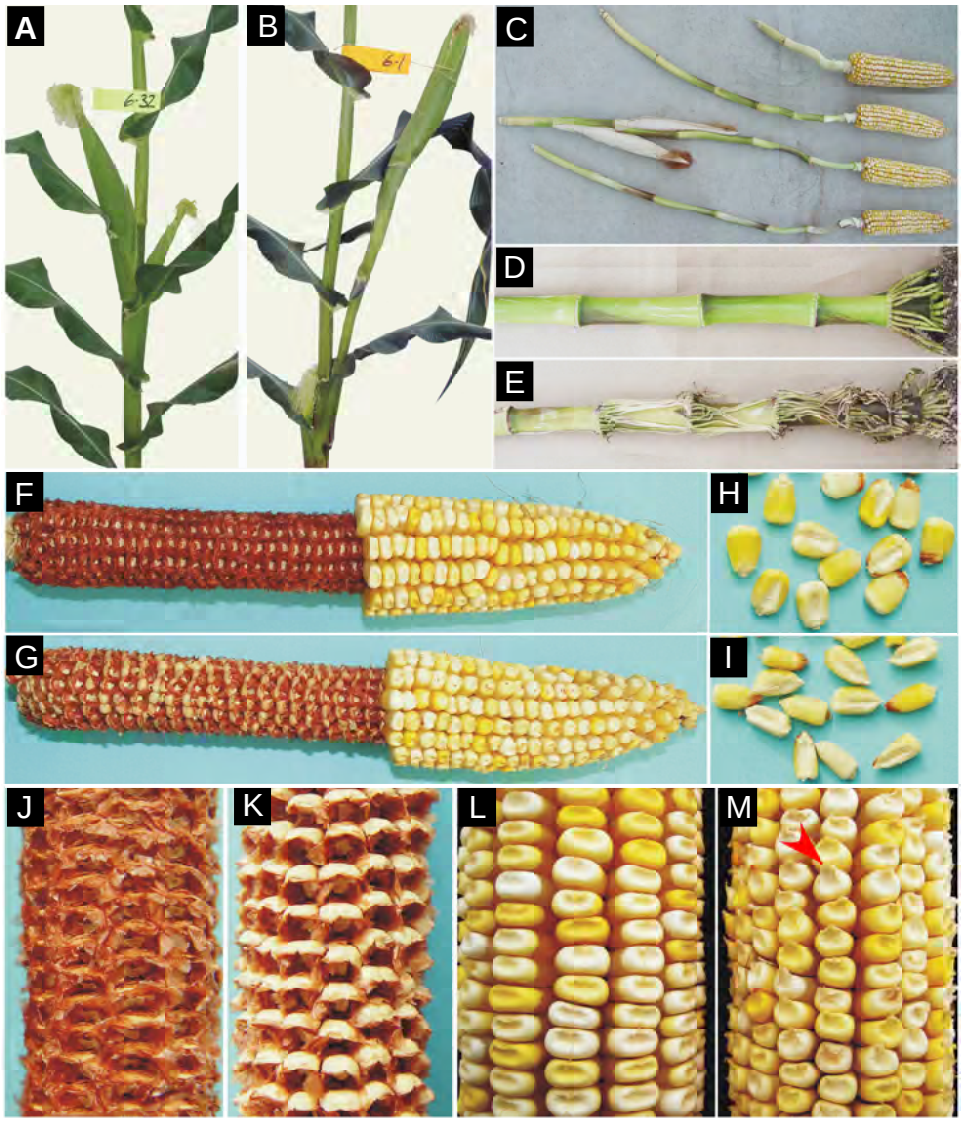
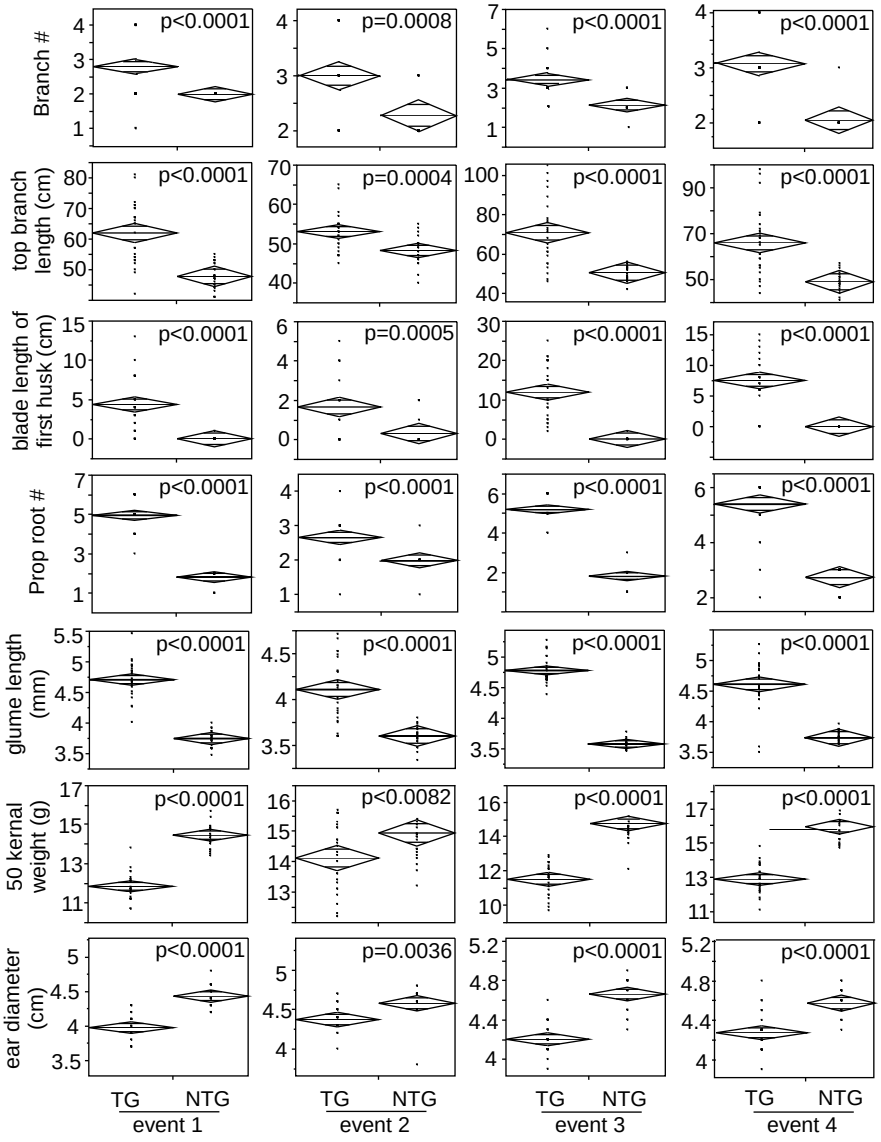


Figure 7


Branch #
p<0.0001
p=0.0008
p<0.0001
p<0.0001
top branch length (cm)
p<0.0001
p=0.0004
p<0.0001
p<0.0001
blade length of first husk (cm)
p<0.0001
p=0.0005
p<0.0001
p<0.0001
Prop root #
p<0.0001
p<0.0001
p<0.0001
p<0.0001
glume length (mm)
p<0.0001
p<0.0001
p<0.0001
p<0.0001
50 kernal weight (g)
p<0.0001
p<0.0082
p<0.0001
p<0.0001
ear diameter (cm)
p<0.0001
p=0.0036
p<0.0001
p<0.0001
TG
NTG
event 1
event 2
event 3
event 4