

Authors: Laura Vann¹, Thomas Kono^{1,2}, Tanja Pyhäjärvi^{1,3}, Matthew B. Hufford^{1,4,6}, and Jeffrey
Ross-Ibarra^{1,5,7}

¹Department of Plant Sciences, University of California Davis, Davis, CA, USA

²Department of Agronomy and Plant Genetics, University of Minnesota Twin Cities,
Minneapolis, MN, USA

³Department of Biology, University of Oulu, Oulu, Finland

⁴Department of Ecology, Evolution, and Organismal Biology, Iowa State University, Ames, Iowa,
USA

⁵Center for Population Biology and Genome Center, University of California Davis, Davis, CA,
USA

⁶Corresponding Author: Matthew B. Hufford; 339A Bessey Hall, Iowa State University, Ames,
IA, USA; phone: 1-515-294-8511; email: mhufford@iastate.edu

⁷Corresponding Author: Jeffrey Ross-Ibarra; 262 Robbins Hall, Mail Stop 4, University of
California, Davis, CA, USA; phone: 1-530-752-1152; email: rossibarra@ucdavis.edu

INTRODUCTION

Domesticated crops and their wild progenitors provide an excellent system in which to study adaptation and genomic changes associated with human-mediated selection (Ross-Ibarra et al., 2007). Plant domestication usually involves a suite of phenotypic changes such as loss of seed shattering and increased fruit or grain size, which are commonly referred to as the ‘domestication syndrome’ (Olsen and Wendel, 2013), and much of the study of domestication has focused on understanding the genetic variation underlying these traits (Olsen and Gross, 2010). Because most domesticates show reduced genetic diversity relative to their wild counterparts, effort has been made to identify agronomically useful variation in crop wild relatives (Flint-Garcia et al., 2009). In some instances, the alleles conferring these beneficial traits are bred into domesticates for crop improvement. For example, *Oryza rufipogon*, the wild progenitor of domesticated rice, has proven useful for the integration of a number of beneficial QTL controlling traits such as grain size and yield into domesticated rice (Kovach and McCouch, 2008). In addition to researching the role of wild alleles in domesticates, researchers have also investigated the role of variation in domesticated taxa in the evolution of feral and weedy populations (Ellstrand et al., 2010). But even though domesticated alleles are often found segregating in wild relatives (Gallavotti et al., 2004; Sigmon and Vollbrecht, 2010), little is known about the ecological role of this variation in natural populations. In this paper we present an ecological genetic analysis of the domestication locus *tb1*—specifically the domesticated haplotype at *tb1*—in natural populations of the wild ancestor of domesticated maize.

Maize (*Zea mays* ssp. *mays*) was domesticated from the teosinte *Zea mays* ssp. *parviglumis* (hereafter, *parviglumis*) roughly 9,000 B.P. in southwest Mexico (Piperno et al., 2009; Matsuoka et al., 2002). Maize and the teosintes are an attractive system in which to study domestication due to the abundance of genetic tools developed for maize and well-characterized domestication loci (Hufford et al., 2012a; Doebley, 2004; Hufford et al., 2012b). Additionally, large, naturally-occurring populations of both *parviglumis* and the highland teosinte *Zea mays* ssp. *mexicana* (hereafter, *mexicana*) can be found throughout Mexico (Wilkes, 1977; Hufford et al., 2013), and genetic diversity of these taxa is estimated to be high (Ross-Ibarra et al., 2009).

Many morphological changes are associated with maize domestication, and understanding the genetic basis of these changes has been a focus of maize research for a number of years (Doebley, 2004). One of the most dramatic changes is found in plant architecture: domesticated maize is characterized by a central stalk with few tillers and lateral branches terminating in a female inflorescence, while teosinte is highly tillered and bears tassels (male inflorescences) at the end of its lateral branches. The *teosinte branched1* (*tb1*) gene, a repressor of organ growth, was identified as a major QTL involved in branching (Doebley et al., 1995) and tillering (Doebley and Stec, 1991) differences between maize and teosinte. A 4.9 kb retrotransposon (*Hopscotch*) insertion into the upstream control region of *tb1* in maize acts to enhance expression of *tb1*, thus repressing lateral organ growth (Doebley et al., 1997; Studer et al., 2011). Dating of the *Hopscotch* retrotransposon suggests that its insertion predates the domestication of maize, leading to the hypothesis that it was segregating as standing variation in populations of teosinte and increased to high frequency in maize due to selection during domestication (Studer et al., 2011). The effects of the *Hopscotch* insertion have been studied in maize (Studer et al., 2011), and analysis of teosinte alleles at *tb1* has identified functionally distinct allelic classes of *tb1* (Studer and Doebley, 2012), but little is known about the role of *tb1* or the *Hopscotch* insertion at this locus in natural populations of teosinte. Previous studies have confirmed the presence of the *Hopscotch* in samples of *parviglumis* and landrace maize (Studer et al., 2011); however, little is known about the frequency with which the *Hopscotch* is segregating in natural populations.

In teosinte and other plants that grow at high population density, individuals detect competition from neighbors via the ratio of red to far-red light. An increase in far-red relative to red light accompanies shading and triggers the shade avoidance syndrome, a suite of physiological and morphological changes such as reduced tillering, increased plant height and early flowering (Kebrom and Brutnell, 2007). The *tb1* locus appears to play an important role in the shade avoidance pathway in *Zea mays* (Lukens and Doebley, 1999) and other grasses (Kebrom and Brutnell, 2007) via changes in expression levels in response to shading. Lukens and Doebley (1999) introgressed the teosinte *tb1* allele into a maize inbred background and noted that under low density conditions plants were highly tillered but that under high density, plants showed significantly reduced tillers and grew taller. Based on these results we hypothesize that the *Hopscotch* (*i.e.*, the domesticated allele) at *tb1* may play a role in the ecology of teosinte, especially in high-density populations. In this study we aim to characterize the distribution of the *Hopscotch* insertion in *parviglumis*, *mexicana*, and landrace maize, and to examine the phenotypic

61 effects of the insertion in *parviglumis*. We use a combination of PCR genotyping for the
62 *Hopscotch* element in our full panel and sequencing of two small regions upstream of *tb1*
63 combined with a larger SNP dataset in a subset of teosinte populations to explore patterns of
64 genetic variation at this locus. Finally, we test for an association between the *Hopscotch* element
65 and tillering phenotypes in samples from a natural population of *parviglumis*.

66 MATERIALS & METHODS

67 **Sampling and genotyping**—We sampled 1,110 individuals from 350 accessions (247 maize
68 landraces, 17 *mexicana* populations, and 86 *parviglumis* populations) and assessed the presence or
69 absence of the *Hopscotch* insertion (Table S1 and Table S2, See Supplemental Materials with the
70 online version of this article). Numbers of individuals sampled ranged from 1-43 for *parviglumis*,
71 1-35 for *mexicana*, and 1-18 for the maize landrace populations. DNA was extracted from leaf
72 tissue using a modified CTAB approach (Doyle and Doyle, 1990; Maloof et al., 1984). We
73 designed primers using PRIMER3 (Rozen and Skaletsky, 2000) implemented in Geneious (Kearse
74 et al., 2012) to amplify the entire *Hopscotch* element, as well as an internal primer allowing us to
75 simultaneously check for possible PCR bias between presence and absence of the *Hopscotch*
76 insertion due to its large size (~5kb). Two PCRs were performed for each individual, one with
77 primers flanking the *Hopscotch* (HopF/HopR) and one with a flanking primer and an internal
78 primer (HopF/HopIntR). Primer sequences are HopF, 5'-TCGTTGATGCTTTGATGGATGG-3';
79 HopR, 5'-AACAGTATGATTTTCATGGGACCG-3'; and HopIntR,
80 5'-CCTCCACCTCTCATGAGATCC-3' (Fig. S1 and Fig. S2, See Supplemental Materials with the
81 online version of this article). Homozygotes show a single band for absence of the element
82 (~300bp) and two bands for presence of the element (~5kb, amplification of the entire element,
83 and ~1.1kb, amplification of part of the element), whereas heterozygotes show all three bands
84 (Table S2, See Supplemental Materials with the online version of this article). Since we developed
85 a PCR protocol for each allele, if only one PCR resolved well, we scored one allele for that
86 individual rather than infer the diploid genotype. We used Phusion High Fidelity Enzyme
87 (Thermo Fisher Scientific Inc., Waltham, Massachusetts, USA) and the following conditions for
88 amplifications: 98°C for 3 min, 30 cycles of 98°C for 15 s, 65°C for 30 s, and 72°C for 3 min 30 s,
89 with a final extension of 72°C for 10 min. PCR products were visualized on a 1% agarose gel and
90 scored for presence/absence of the *Hopscotch* based on band size.

Genotyping analysis—To calculate differentiation between populations (F_{ST}) and subspecies (F_{CT}) we used HierFstat (Goudet, 2005). These analyses only included populations ($n = 32$) in which eight or more chromosomes were sampled. To test the hypothesis that the *Hopscotch* insertion may be adaptive under certain environmental conditions, we looked for significant associations between *Hopscotch* frequency and environmental variables using the software BayEnv (Coop et al., 2010). BayEnv creates a covariance matrix of relatedness between populations and then tests a null model that allele frequencies in populations are determined by the covariance matrix of relatedness alone against the alternative model that allele frequencies are determined by a combination of the covariance matrix and an environmental variable, producing a posterior probability (*i.e.*, Bayes Factor; Coop et al. 2010). We used teosinte (*ssp. parviglumis* and *ssp. mexicana*) genotyping and covariance data from Pyhäjärvi et al. (2013) for BayEnv, with the *Hopscotch* insertion coded as an additional biallelic marker. SNP data from Pyhäjärvi et al. (2013) were obtained using the MaizeSNP50 BeadChip and Infinium HD Assay (Illumina, San Diego, CA, USA). Environmental data were obtained from www.worldclim.org and soil data were downloaded from the Harmonized World Soil Database (FAO/IIASA/ISRIC/ISSCAS/JRC, 2012) at www.harvestchoice.org. Environmental data represent average values for the last several decades (climatic data) or are likely stable over time (soil data) and therefore represent conditions important for local adaptation of our samples. Information from these data sets was summarized by principle component analysis following Pyhäjärvi et al. (2013).

Sequencing—In addition to genotyping, we chose a subset of *parviglumis* individuals for sequencing. We chose twelve individuals from each of four populations from Jalisco state, Mexico (San Lorenzo, La Mesa, Ejutla A, and Ejutla B). For amplification and sequencing, we selected two regions approximately 600bp in size from within the 5' UTR of *tb1* (Region 1) and from 1,235bp upstream of the start of the *Hopscotch* (66,169bp upstream from the start of the *tb1* ORF; Region 2). We designed the following primers using PRIMER3 (Rozen and Skaletsky, 2000): for the 5' UTR, 5'-GGATAATGTGCACCAGGTGT-3' and 5'-GCGTGCTAGAGACACYTGTGCT-3'; for the 66kb upstream region, 5'-TGTCCTCGCCGCAACTC-3' and 5'-TGTACGCCCGCCCTCATCA-3' (Table S1, See Supplemental Materials with the online version of this article). We used Taq polymerase (New England Biolabs Inc., Ipswich, Massachusetts, USA) and the following thermal cycler conditions to amplify fragments: 94°C for 3 min, 30 cycles of 92°C for 40 s, annealing for 1 min, 72°C for 40 s, and a final 10 min extension at 72°C. Annealing temperatures for Region 1 and Region 2 were 59.7°C and 58.8°C, respectively. To

clean excess primer and dNTPs we added two units of Exonuclease1 and 2.5 units of Antarctic Phosphatase to 8.0 μ L of amplification product. This mix was placed on a thermal cycler with the following program: 37°C for 30 min, 80°C for 15 min, and a final cool-down step to 4°C.

We cloned cleaned fragments into a TOPO-TA vector (Life Technologies, Grand Island, New York, USA) using OneShot TOP10 chemically competent *E. coli* cells, with an extended ligation time of 30 min for a complex target fragment. We plated cells on LB agar plates containing kanamycin, and screened colonies using vector primers M13 Forward and M13 Reverse under the following conditions: 96°C for 5 min; then 35 cycles at 96°C for 30 s, 53°C for 30 s, 72°C for 2 min; and a final extension at 72°C for 4 min. We visualized amplification products for incorporation of our insert on a 1% agarose TAE gel.

Amplification products with successful incorporation of our insert were cleaned using Exonuclease 1 and Antarctic Phosphatase following the procedures detailed above, and sequenced with vector primers M13 Forward and M13 Reverse using Sanger sequencing at the College of Agriculture and Environmental Sciences (CAES) sequencing center at UC Davis. We aligned and trimmed primer sequences from resulting sequences using the software Geneious (Kearse et al., 2012). Following alignment, we verified singleton SNPs by sequencing an additional one to four colonies from each clone. If the singleton was not present in these additional sequences it was considered an amplification or cloning error, and we replaced the base with the base of the additional sequences. If the singleton appeared in at least one of the additional sequences we considered it a real variant and kept it for further analyses.

Sequence analysis—For population genetic analyses of sequenced Region 1 and sequenced Region 2 we used the Libsequence package (Thornton, 2003) to calculate pairwise F_{ST} between populations and to calculate standard diversity statistics (number of haplotypes, haplotype diversity, Watterson’s estimator $\hat{\theta}_W$, pairwise nucleotide diversity $\hat{\theta}_\pi$, and Tajima’s D). To produce a visual representation of differentiation between sequences and examine patterns in sequence clustering by *Hopscotch* genotype we used Phylip (<http://evolution.genetics.washington.edu/phylip.html>) to create neighbor-joining trees with bootstrap-supported nodes (100 repetitions). For creation of trees we also included homologous sequence data from Maize HapMapV2 (Chia et al., 2012) for teosinte inbred lines (TILs), some of which are known to be homozygous for the *Hopscotch* insertion (TIL03, TIL17, TIL09), as well as 59 lines of domesticated maize.

Introgression analysis—In order to assess patterns of linkage disequilibrium (LD) around

the *Hopscotch* element in the context of chromosomal patterns of LD we used Tassel (Bradbury et al., 2007) and calculated LD between SNPs across chromosome 1 using previously published data from twelve plants each of the Ejutla A (EjuA), Ejutla B (EjuB), San Lorenzo (SLO), and La Mesa (MSA) populations (Pyhäjärvi et al., 2013). We chose these populations because we had both genotyping data for the *Hopscotch* as well as chromosome-wide SNP data for chromosome 1. For each population we filtered the initial set of 5,897 SNPs on chromosome 1 to accept only SNPs with a minor allele frequency of at least 0.1, resulting in 1,671, 3,023, 3,122, and 2,167 SNPs for SLO, EjuB, EjuA, and MSA, respectively. We then used Tassel (Bradbury et al., 2007) to calculate linkage disequilibrium (r^2) across chromosome 1 for each population.

We examined evidence of introgression on chromosome 1 in these same four populations (EjuA, EjuB, MSA, SLO) using STRUCTURE (Falush et al., 2003) and phased data from Pyhäjärvi et al. (2013), combined with the corresponding SNP data from a diverse panel of 282 maize lines (Cook et al., 2012). SNPs were anchored in a modified version of the IBM genetic map (Gerke et al., 2013). Since STRUCTURE does not account for LD due to physical linkage we created haplotype blocks using a custom Perl script from Hufford et al. (2013, code available at <http://dx.doi.org/10.6084/m9.figshare.1165577>). In maize, LD decays over an average distance of 5500bp (Chia et al., 2012); because LD decay is even more rapid in teosinte (Chia et al., 2012) we used a conservative haplotype block size of 5kb. We ran STRUCTURE at K=2 under the linkage model, with the assumption being that individuals fall into either a maize or teosinte cluster, performing three replicates with an MCMC burn-in of 10,000 steps and 50,000 steps post burn-in.

Phenotyping of *parviglumis*—To investigate the phenotypic effects of the *Hopscotch* insertion in teosinte we conducted a phenotyping trial in which we germinated 250 seeds of *parviglumis* collected in Jalisco state, Mexico (population San Lorenzo; Hufford 2010) where the *Hopscotch* insertion is segregating at highest frequency (0.44) in our initial genotyping sample set. In order to maximize the likelihood of finding the *Hopscotch* in our association population we selected seeds from sites within the population where genotyped individuals were homozygous or heterozygous for the insertion. We chose between 10-13 seeds from each of 23 sampling sites. We treated seeds with Captan fungicide (Southern Agricultural Insecticides Inc., Palmetto, Florida, USA) and germinated them in petri dishes with filter paper. Following germination, 206 successful germinations were planted into one-gallon pots with potting soil and randomly spaced one foot apart on greenhouse benches. Plants were watered three times a day with an automatic

187 drip containing 10-20-10 fertilizer, which was supplemented with hand watering on extremely hot
188 and dry days.

189 Starting on day 15, we measured tillering index as the ratio of the sum of tiller lengths to the
190 height of the plant (Briggs et al., 2007). Following initial measurements, we phenotyped plants for
191 tillering index every 5 days through day 40, and then on day 50 and day 60. On day 65 we
192 measured culm diameter between the third and fourth nodes of each plant. Following
193 phenotyping we extracted DNA from all plants using a modified SDS extraction protocol. We
194 genotyped individuals for the *Hopscotch* insertion following the PCR protocols listed above.

195 Tillering index data for each genotypic class did not meet the criteria for a repeated measures
196 ANOVA, so we transformed the data with a Box-Cox transformation ($\lambda = 0$) in the Car Package
197 for R (Fox and Weisberg, 2011) to improve the normality and homogeneity of variance among
198 genotype classes. We analyzed relationships between genotype and tillering index and tiller
199 number using a repeated measures ANOVA through a general linear model function implemented
200 in SAS v.9.3 (SAS Institute Inc., Cary, NC, USA). Additionally, in order to compare any
201 association between *Hopscotch* genotype and tillering and associations at other presumably
202 unrelated traits, we performed an ANOVA between culm diameter and genotype using the same
203 general linear model in SAS. Culm diameter is not believed to be correlated with tillering index
204 or variation at *tb1* and is used as our independent trait for phenotyping analyses. SAS code used
205 for analysis is available at <http://dx.doi.org/10.6084/m9.figshare.1166630>.

206 RESULTS

207 **Genotyping**—The genotype at the *Hopscotch* insertion was confirmed with two PCRs for
208 837 individuals of the 1,100 screened. Among the 247 maize landrace accessions genotyped, all
209 but eight were homozygous for the presence of the insertion (Table S1 and Table S2, See
210 Supplemental Materials with the online version of this article). Within our *parviglumis* and
211 *mexicana* samples we found the *Hopscotch* insertion segregating in 37 ($n = 86$) and four ($n = 17$)
212 populations, respectively, and at highest frequency within populations in the states of Jalisco,
213 Colima, and Michoacán in central-western Mexico (Fig. 1). Using our *Hopscotch* genotyping, we
214 calculated differentiation between populations (F_{ST}) and subspecies (F_{CT}) for populations in
215 which we sampled sixteen or more chromosomes. We found that $F_{CT} = 0$, and levels of F_{ST}
216 among populations within each subspecies (0.22) and among all populations (0.23) are similar to
217 genome-wide estimates from previous studies (Pyhäjärvi et al. 2013; Table 1). Although we found

large variation in *Hopscotch* allele frequency among our populations, BayEnv analysis did not indicate a correlation between the *Hopscotch* insertion and environmental variables (all Bayes Factors < 1).

Sequencing—To investigate patterns of sequence diversity and linkage disequilibrium (LD) in the *tb1* region and any evidence of selection on this locus, we sequenced two small (<1kb) regions upstream of the *tb1* ORF in four populations. After alignment and singleton checking we recovered 48 and 40 segregating sites for the 5' UTR region (Region 1) and the 66kb upstream region (Region 2), respectively. For Region 1, Ejutla A has the highest values of haplotype diversity and θ_π , while Ejutla B and La Mesa have comparable values of these summary statistics, and San Lorenzo has much lower values. Additionally, Tajima's D is strongly negative in the two Ejutla populations and La Mesa, but is closer to zero in San Lorenzo (Table 2, Table S2, See Supplemental Materials with the online version of this article). For Region 2, haplotype diversity and θ_π , are similar for Ejutla A and Ejutla B, while La Mesa and San Lorenzo have slightly lower values for these statistics (Table 2). Tajima's D is positive in all populations except La Mesa, indicating an excess of low frequency variants in this population (Table 2). Pairwise values of F_{ST} within population pairs Ejutla A/Ejutla B and San Lorenzo/La Mesa are close to zero for both sequenced regions as well as for the *Hopscotch*, while they are high for other population pairs (Table 1). Neighbor joining trees of our sequence data and data from the teosinte inbred lines (TILs; data from Maize HapMapV2, Chia et al. 2012) do not reveal any clear clustering pattern with respect to population or *Hopscotch* genotype (Fig. S3, See Supplemental Materials with the online version of this article); individuals within our sample that have the *Hopscotch* insertion do not group with the teosinte inbred lines or domesticated maize that have the *Hopscotch* insertion.

Evidence of introgression—The highest frequency of the *Hopscotch* insertion in teosinte was found in *parviglumis* sympatric with cultivated maize. Our initial hypothesis was that the high frequency of the *Hopscotch* element in these populations could be attributed to introgression from maize into teosinte. To investigate this possibility we examined overall patterns of linkage disequilibrium across chromosome 1 and specifically in the *tb1* region. If the *Hopscotch* is found in these populations due to recent introgression from maize we would expect to find large blocks of linked markers near this element. We find no evidence of elevated linkage disequilibrium between the *Hopscotch* and SNPs surrounding the *tb1* region in our resequenced populations (Fig. 2), and r^2 in the *tb1* region does not differ significantly between populations with (average r^2 of 0.085) and without (average $r^2 = 0.082$) the *Hopscotch* insertion. In fact, average r^2 is lower

250 in the *tb1* region ($r^2 = 0.056$) than across the rest of chromosome 1 ($r^2 = 0.083$; Table 3).

251 The lack of clustering of *Hopscotch* genotypes in our NJ tree as well as the lack of LD around
252 *tb1* do not support the hypothesis that the *Hopscotch* insertion in these populations of *parviglumis*
253 is the result of recent introgression. However, to further explore this hypothesis we performed a
254 STRUCTURE analysis using Illumina MaizeSNP50 data from four of our *parviglumis* populations
255 (EjuA, EjuB, MSA, and SLO) (Pyhäjärvi et al., 2013) and the maize 282 diversity panel (Cook
256 et al., 2012). The linkage model implemented in STRUCTURE can be used to identify ancestry of
257 blocks of linked variants which would arise as the result of recent admixture between populations.
258 If the *Hopscotch* insertion is present in populations of *parviglumis* as a result of recent admixture
259 with domesticated maize, we would expect the insertion and linked variants in surrounding sites
260 to be assigned to the “maize” cluster in our STRUCTURE runs, not the “teosinte” cluster. In all
261 runs, assignment to maize in the *tb1* region across all four *parviglumis* populations is low (average
262 0.017) and much below the chromosome-wide average (0.20; Table 4; Fig. 3).

263 **Phenotyping**—To assess the contribution of *tb1* to phenotypic variation in tillering in a
264 natural population, we grew plants from seed sampled from the San Lorenzo population of
265 *parviglumis*, which had a high mean frequency (0.44) of the *Hopscotch* insertion based on our
266 initial genotyping. We measured tiller number and tillering index, the ratio of the sum of tiller
267 lengths to plant height, for 206 plants from within the San Lorenzo population, and genotyped
268 plants for the *Hopscotch* insertion. We also measured culm diameter, a phenotype that differs
269 between maize and teosinte but has not been shown to be affected by the *Hopscotch* insertion
270 (Briggs et al., 2007). Culm diameter is meant to be an independent trait against which we can
271 compare patterns of tillering index x *Hopscotch* genotype data. If tillering index in *parviglumis* is
272 affected by the *Hopscotch* insertion, the expectation is that patterns of tillering index data will
273 have a significant correlation with *Hopscotch* genotype, whereas we should find no significant
274 correlation between culm diameter and *Hopscotch* genotype. Phenotypic data are available at
275 <http://dx.doi.org/10.6084/m9.figshare.776926>. Our plantings produced 82 homozygotes
276 for the *Hopscotch* insertion at *tb1*, 104 heterozygotes, and 20 homozygotes lacking the insertion;
277 these numbers do not deviate from expectations of Hardy-Weinberg equilibrium. After
278 performing a repeated measures ANOVA between our transformed tillering index data and
279 *Hopscotch* genotype we find no significant correlation between genotype at the *Hopscotch*
280 insertion and tillering index (Fig. 4), tiller number, or culm diameter. Only on day 40 did we
281 observe a weak but statistically insignificant ($r^2 = 0.02$, $p = 0.0848$) correlation between tillering

index and the *Hopscotch* genotype, although in the opposite direction of that expected, with homozygotes for the insertion showing a higher tillering index.

DISCUSSION

Adaptation occurs due to selection on standing variation or *de novo* mutations. Adaptation from standing variation has been well-described in a number of systems; for example, selection for lactose tolerance in humans (Plantinga et al., 2012; Tishkoff et al., 2007), variation at the *Eda* locus in three-spined stickleback (Kitano et al., 2008; Colosimo et al., 2005), and pupal diapause in the Apple Maggot fly (Feder et al., 2003). Although the adaptive role of standing variation has been described in many systems, its importance in domestication is not as well studied.

In maize, alleles at domestication loci (*RAMOSA1*, Sigmon and Vollbrecht 2010; *barren stalk1*, Gallavotti et al. 2004; and *grassy tillers1*, Whipple et al. 2011) are thought to have been selected from standing variation, suggesting that diversity already present in teosinte may have played an important role in maize domestication. The *teosinte branched1* gene is one of the best characterized domestication loci, and, while previous studies have suggested that differences in plant architecture between maize and teosinte are a result of selection on standing variation at this locus (Clark et al., 2006; Studer et al., 2011), much remains to be discovered regarding natural variation at this locus and its ecological role in teosinte.

Studer et al. (2011) genotyped 90 accessions of teosinte (inbred and outbred), providing the first evidence that the *Hopscotch* insertion is segregating in teosinte. Given that the *Hopscotch* insertion has been estimated to predate the domestication of maize, it is not surprising that it can be found segregating in populations of teosinte. However, by widely sampling across teosinte populations our study provides greater insight into the distribution and prevalence of the *Hopscotch* in teosinte. While our findings are consistent with Studer et al. (2011) in that we identify the *Hopscotch* allele segregating in teosinte, we find it at higher frequency than previously suggested. The *Hopscotch* allele is more prevalent in *parviglumis* than in *mexicana* in our sample (Table S1, See Supplemental Materials with the online version of this article), suggesting a different history of the allele amongst teosinte subspecies. Moreover, many of our *parviglumis* populations with a high frequency of the *Hopscotch* allele fall in the Jalisco cluster identified by Fukunaga et al. (2005), and further distinguish this region from the Balsas River Basin where maize was domesticated (Matsuoka et al., 2002). Potential explanations for the high frequency of the *Hopscotch* element in *parviglumis* from the Jalisco cluster include gene flow from

313 maize, genetic drift, and natural selection.

314 While gene flow from crops into their wild relatives is well-known, (Ellstrand et al., 1999;
315 Zhang et al., 2009; Thurber et al., 2010; Baack et al., 2008; Hubner et al., 2012; Wilkes, 1977; van
316 Heerwaarden et al., 2011; Barrett, 1983), our results do not suggest introgression from maize at
317 the *tb1* locus, and are more consistent with Hufford et al. (2013) who found resistance to
318 introgression from maize into *mexicana* around domestication loci. Clustering in our NJ trees
319 does not reflect the pattern expected if maize alleles at the *tb1* locus had introgressed into
320 populations of teosinte. Moreover, there is no signature of elevated LD in the *tb1* region relative
321 to the rest of chromosome 1, and Bayesian assignment to a maize cluster in this region is both low
322 and below the chromosome-wide average (Fig. 3, Table 4). Together, these data point to an
323 explanation other than recent introgression for the high observed frequency of *Hopscotch* in the
324 Jalisco cluster of our *parviglumis* populations.

325 Although recent introgression seems unlikely, we cannot rule out ancient introgression as an
326 explanation for the presence of the *Hopscotch* in these populations. If the *Hopscotch* allele was
327 introgressed in the distant past, recombination may have broken up LD, a process that would be
328 consistent with our data. We find this scenario less plausible, however, as there is no reason why
329 gene flow should have been high in the past but absent in present-day sympatric populations. In
330 fact, early generation maize-teosinte hybrids are common in these populations today (MB
331 Hufford, pers. observation), and genetic data support ongoing gene flow between domesticated
332 maize and both *mexicana* and *parviglumis* in a number of sympatric populations (Hufford et al.,
333 2013; Ellstrand et al., 2007; van Heerwaarden et al., 2011; Warburton et al., 2011).

334 Remaining explanations for differential frequencies of the *Hopscotch* among teosinte
335 populations include both genetic drift and natural selection. Previous studies using both SSRs
336 and genome-wide SNP data have found evidence for a population bottleneck in the San Lorenzo
337 population (Hufford, 2010; Pyhäjärvi et al., 2013), and the lower levels of sequence diversity in
338 this population in the 5' UTR (Region 1) coupled with more positive values of Tajima's D are
339 consistent with these earlier findings. Such population bottlenecks can exaggerate the effects of
340 genetic drift through which the *Hopscotch* allele may have risen to high frequency entirely by
341 chance. A bottleneck in San Lorenzo, however, does not explain the high frequency of the
342 *Hopscotch* in multiple populations in the Jalisco cluster. Moreover, available information on
343 diversity and population structure among Jaliscan populations (Hufford, 2010; Pyhäjärvi et al.,
344 2013) is not suggestive of recent colonization or other demographic events that would predict a

high frequency of the allele across populations. Finally, diversity values in the 5' UTR of *tb1* are suggestive of natural selection acting upon the gene in populations of *parviglumis*. Overall nucleotide diversity is 76% less than seen in the sequences from the 66kb upstream region, and Tajima's D is considerably lower and consistently negative across populations (Table 2). In fact, values of Tajima's D in the 5' UTR are toward the extreme negative end of the distribution of this statistic previously calculated across loci sequenced in *parviglumis* (Wright et al., 2005; Moeller et al., 2007). Though not definitive, these results are consistent with the action of selection on the upstream region of *tb1*, perhaps suggesting an ecological role for the gene in Jaliscoan populations of *parviglumis*. Finally, while these results are consistent with selection at the *tb1* locus in teosinte, they do not confirm selection specifically on the *Hopscotch* insertion at this locus.

Significant effects of the *Hopscotch* insertion on lateral branch length, number of cupules, and tillering index in domesticated maize have been well documented (Studer et al., 2011). Weber et al. (2007) described significant phenotypic associations between markers in and around *tb1* and lateral branch length and female ear length in a sample from 74 natural populations of *parviglumis* (Weber et al., 2007); however, these data did not include markers from the *Hopscotch* region 66kb upstream of *tb1*. Our study is the first to explicitly examine the phenotypic effects of the *Hopscotch* insertion across a wide collection of individuals sampled from natural populations of teosinte. We have found no significant effect of the *Hopscotch* insertion on tillering index or tiller number, a result that is discordant with its clear phenotypic effects in maize. One interpretation of this result would be that the *Hopscotch* controls tillering in maize (Studer et al., 2011), but tillering in teosinte is affected by variation at other loci. Consistent with this interpretation, *tb1* is thought to be part of a complex pathway controlling branching, tillering and other phenotypic traits (Kebrom and Brutnell, 2007; Clark et al., 2006).

A recent study by Studer and Doebley (2012) examined variation across traits in a three-taxa allelic series at the *tb1* locus. Studer and Doebley (2012) introgressed nine unique teosinte *tb1* segments (one from *Zea diploperennis*, and four each from *mexicana* and *parviglumis*) into an inbred maize background and investigated their phenotypic effects. Their findings suggest that different teosinte *tb1* segments contain distinct genetic factors that differentially contribute to tillering in teosinte. Furthermore, Studer and Doebley (2012) found significant variation in tillering within their control maize lines, concluding that variation in *tb1* alone cannot account for the differences in tillering, and that other genetic factors must also contribute to differences in tillering in teosinte. Clues to the identity of these loci may be found in QTL studies that have

377 identified loci controlling branching architecture (*e.g.*, Doebley and Stec 1991, 1993). Many of
378 these loci (*grassy tillers*, *gt1*; *tassel-replaces-upper-ears1*, *tru1*; *terminal ear1*, *te1*) have been
379 shown to interact with *tb1* (Whipple et al., 2011; Li, 2012), and both *tru1* and *te1* affect the
380 same phenotypic traits as *tb1* (Doebley et al., 1995). *tru1*, for example, has been shown to act
381 either epistatically or downstream of *tb1*, affecting both branching architecture (decreased apical
382 dominance) and tassel phenotypes (shortened tassel and shank length and reduced tassel number;
383 Li 2012). Variation in these additional loci may have affected tillering in our collections and
384 contributed to the lack of correlation we see between *Hopscotch* genotype and tillering.

385 CONCLUSIONS

386 In conclusion, our findings demonstrate that the *Hopscotch* allele is widespread in populations
387 of *parviglumis* and *mexicana* and occasionally at high allele frequencies. Analysis of linkage using
388 SNPs from across chromosome 1 does not suggest that the *Hopscotch* allele is present in these
389 populations due to recent introgression, and it seems unlikely that the insertion would have
390 drifted to high frequency in multiple populations. We do, however, find preliminary evidence of
391 selection on the *tb1* locus in *parviglumis*. Coupled with our observation of high frequency of the
392 *Hopscotch* insertion in a number of populations, this suggests that the locus—and potentially the
393 domestication allele at this locus—may play an ecological role in teosinte.

394 In contrast to domesticated maize, the *Hopscotch* insertion does not appear to have a large
395 effect on tillering in a diverse sample of *parviglumis* from a natural population and the phenotypic
396 consequences of variation at *tb1* thus remain unclear. Future studies should examine expression
397 levels of *tb1* in teosinte with and without the *Hopscotch* insertion and further characterize the
398 effects of additional loci involved in branching architecture (*e.g.* *gt1*, *tru1*, and *te1*). These data,
399 in conjunction with more exhaustive phenotyping, should help to further clarify the ecological
400 significance of the domesticated *tb1* allele in natural populations of teosinte.

401 Acknowledgements

402 The authors thank Graham Coop for helpful discussion.

REFERENCES

- BAACK, E., Y. SAPIR, M. CHAPMAN, J. BURKE, AND L. RIESEBERG. 2008. Selection on domestication traits and quantitative trait loci in crop-wild sunflower hybrids. *Mol Ecol* 17: 666–677.
- BARRETT, S. 1983. Crop mimicry in weeds. *Econ Bot* 37: 255–282.
- BRADBURY, P., Z. ZHANG, D. KROON, T. CASSTEVEN, Y. RAMDOSS, AND E. BUCKLER. 2007. Tassel: software for association mapping of complex traits in diverse samples. *Bioinformatics* 23: 2633–2635.
- BRIGGS, W., M. McMULLEN, B. GAUT, AND J. DOEBLEY. 2007. Linkage mapping of domestication loci in a large maize-teosinte backcross resource. *Genetics* 177: 1915–1928.
- CHIA, J., C. SONG, P. BRADBURY, D. COSTICH, N. DE, LEON, J. DOEBLEY, R. ELSHIRE, B. GAUT, L. GELLER, J. GLAUBITZ, M. GORE, K. GUILL, J. HOLLAND, M. HUFFORD, J. LAI, M. LI, X. LIU, Y. LU, R. McCOMBIE, R. NELSON, J. POLAND, B. PRASANNA, T. PYHÄJÄRVI, T. RONG, R. SEKHON, Q. SUN, M. TENAILLON, F. TIAN, J. WANG, X. XU, Z. ZHANG, S. KAEPPLE, J. ROSS-IBARRA, M. McMULLEN, E. BUCKLER, G. ZHANG, Y. XU, AND D. WARE. 2012. Maize hapmap2 identifies extant variation from a genome in flux. *Nat Genet* 44: 803–U238.
- CLARK, R., T. WAGLER, P. QUIJADA, AND J. DOEBLEY. 2006. A distant upstream enhancer at the maize domestication gene *tb1* has pleiotropic effects on plant and inflorescent architecture. *Nat Genet* 38: 594–597.
- COLOSIMO, P., K. HOSEMAN, S. BALABHADRA, G. VILLARREAL, M. DICKSON, J. GRIMWOOD, J. SCHMUTZ, R. MYERS, D. SCHLUTER, AND D. KINGSLEY. 2005. Widespread parallel evolution in sticklebacks by repeated fixation of ectodysplasin alleles. *Science* 307: 1928–1933.
- COOK, J., M. McMULLEN, J. HOLLAND, F. TIAN, P. BRADBURY, J. ROSS-IBARRA, E. BUCKLER, AND S. FLINT-GARCIA. 2012. Genetic architecture of maize kernel composition in the nested association mapping and inbred association panels. *Plant Physiol* 158: 824–834.
- COOP, G., D. WITONSKY, A. DI, RIENZO, AND J. PRITCHARD. 2010. Using environmental correlations to identify loci underlying local adaptation. *Genetics* 185: 1411–1423.

- 432 DOEBLEY, J. 2004. The genetics of maize evolution. *Annu Rev Genet* 38: 37–59.
- 433 DOEBLEY, J. AND A. STEC. 1991. Genetic-analysis of the morphological differences between
434 maize and teosinte. *Genetics* 129: 285–295.
- 435 DOEBLEY, J. AND A. STEC. 1993. Inheritance of the morphological differences between maize
436 and teosinte: Comparison of results for two F₂ populations. *Genetics* 134: 559–570.
- 437 DOEBLEY, J., A. STEC, AND C. GUSTUS. 1995. *teosinte branched1* and the origin of maize:
438 Evidence for epistasis and the evolution of dominance. *Genetics* 141: 333–346.
- 439 DOEBLEY, J., A. STEC, AND L. HUBBARD. 1997. The evolution of apical dominance in maize.
440 *Nature* 386: 485–488.
- 441 DOYLE, J. AND J. DOYLE. 1990. A rapid total dna preparation procedure for small quantities of
442 fresh tissue. *Phytochemical Bulletin* 19: 11–15.
- 443 ELLSTRAND, N., L. GARNER, S. HEGDE, R. GUADAGNUOLO, AND L. BLANCAS. 2007.
444 Spontaneous hybridization between maize and teosinte. *Journal of Heredity* 98: 183–187.
- 445 ELLSTRAND, N., H. PRENTICE, AND J. HANCOCK. 1999. Gene flow and introgression from
446 domesticated plants into their wild relatives. *Annu Rev Ecol Syst* 30: 539–563.
- 447 ELLSTRAND, N. C., S. M. HEREDIA, J. A. LEAK-GARCIA, J. M. HERATY, J. C. BURGER,
448 L. YAO, S. NOHZADEH-MALAKSHAH, AND C. E. RIDLEY. 2010. Crops gone wild: evolution of
449 weeds and invasives from domesticated ancestors. *Evolutionary Applications* 3: 494–504.
- 450 FALUSH, D., M. STEPHENS, AND J. PRITCHARD. 2003. Inference of population structure using
451 multilocus genotype data: Linked loci and correlated allele frequencies. *Genetics* 164:
452 1567–1587.
- 453 FAO/IIASA/ISRIC/ISSCAS/JRC. 2012. Harmonized World Soil Database, version 1.2. FAO,
454 Rome, Italy and IIASA, Laxenburg, Austria.
- 455 FEDER, J., S. BERLOCHER, J. ROETHELE, H. DAMBROSKI, J. SMITH, W. PERRY,
456 V. GAVRILOVIC, K. FILCHAK, J. RULL, AND M. ALUJA. 2003. Allopatric genetic origins for
457 sympatric host-plant shifts and race formation in *rhagoletis*. *P Natl Acad Sci Usa* 100:
458 10314–10319.

459 FLINT-GARCIA, S. A., A. L. BODNAR, AND M. P. SCOTT. 2009. Wide variability in kernel
 460 composition, seed characteristics, and zein profiles among diverse maize inbreds, landraces, and
 461 teosinte. *Theoretical and applied genetics* 119: 1129–1142.

462 FOX, J. AND S. WEISBERG. 2011. An R Companion to Applied Regression, vol. Second Edition.
 463 Sage, Thousand Oaks, CA.

464 FUKUNAGA, K., T. NUSSBAUM-WAGLER, B. LI, Q. ZHAO, Y. VIGOUROUX, M. FALLER,
 465 K. BOMBLIES, L. LUKENS, AND J. DOEBLEY. 2005. Genetic diversity and population
 466 structure of teosinte. *Genetics* 169: 2241–2254.

467 GALLAVOTTI, A., Q. ZHAO, J. KYOZUKA, R. MEELEY, M. RITTER, J. DOEBLEY, M. PE, AND
 468 R. SCHMIDT. 2004. The role of barren stalk1 in the architecture of maize. *Nature* 432: 630–635.

469 GERKE, J., J. EDWARDS, G. KE, J. ROSS-IBARRA, AND M. McMULLEN. 2013. The genomic
 470 impacts of drift and selection for hybrid performance in maize. *arXiv* 1307.7313.

471 GOUDET, J. 2005. Hierfstat, a package for r to compute and test hierarchical f-statistics. *Mol*
 472 *Ecol Notes* 5: 184–186.

473 HUBNER, S., T. GUNTHER, A. FLAVELL, E. FRIDMAN, A. GRANER, A. KOROL, AND
 474 K. SCHMID. 2012. Islands and streams: clusters and gene flow in wild barley populations from
 475 the levant. *Mol Ecol* 21: 1115–1129.

476 HUFFORD, M. 2010. Genetic and ecological approaches to guide conservation of teosinte (zea
 477 mays ssp. parviglumis), the wild progenitor of maize. *PhD Dissertation* : 130pp.

478 HUFFORD, M., P. BILINSKI, T. PYHÄJÄRVI, AND J. ROSS-IBARRA. 2012a. Teosinte as a model
 479 system for population and ecological genomics. *Trends in Genetics* 12: 606–615.

480 HUFFORD, M., P. LUBINSKY, T. PYHÄJÄRVI, M. DEVENGENDO, N. ELLSTRAND, AND
 481 J. ROSS-IBARRA. 2013. The genomic signature of crop-wild introgression in maize. *PLoS*
 482 *Genetics* 9: e1003477.

483 HUFFORD, M., X. XU, J. VAN, HEERWAARDEN, T. PYHÄJÄRVI, J. CHIA, R. CARTWRIGHT,
 484 R. ELSHIRE, J. GLAUBITZ, K. GUILL, S. KAEPLER, J. LAI, P. MORRELL, L. SHANNON,
 485 C. SONG, N. SPRINGER, R. SWANSON-WAGNER, P. TIFFIN, J. WANG, G. ZHANG,
 486 J. DOEBLEY, M. McMULLEN, D. WARE, E. BUCKLER, S. YANG, AND J. ROSS-IBARRA.

487 2012b. Comparative population genomics of maize domestication and improvement. *Nat Genet*
 488 44: 808–U118.

489 KEARSE, M., R. MOIR, A. WILSON, S. STONES-HAVAS, M. CHEUNG, S. STURROCK,
 490 S. BUXTON, A. COOPER, S. MARKOWITZ, C. DURAN, T. THIERER, B. ASHTON,
 491 P. MEINTJES, AND A. DRUMMOND. 2012. Geneious basic: An integrated and extendable
 492 desktop software platform for the organization and analysis of sequence data. *Bioinformatics*
 493 28: 1647–1649.

494 KEBROM, T. AND T. BRUTNELL. 2007. The molecular analysis of the shade avoidance syndrome
 495 in the grasses has begun. *Journal of Experimental Botany* 58: 3079–3089.

496 KITANO, J., D. BOLNICK, D. BEAUCHAMP, M. MAZUR, S. MORI, T. NAKANO, AND
 497 C. PEICHEL. 2008. Reverse evolution of armor plates in the threespine stickleback. *Curr Biol*
 498 18: 769–774.

499 KOVACH, M. AND S. MCCOUCH. 2008. Leveraging natural diversity: back through the
 500 bottleneck. *Genome studies and Molecular Genetics* 11: 193–200.

501 LI, W. 2012. Tassels replace upper ears1 encodes a putative transcription factor that regulates
 502 maize shoot architecture by multiple pathways. *PhD Dissertation* : 122.

503 LUKENS, L. AND J. DOEBLEY. 1999. Epistatic and environmental interactions for quantitative
 504 trait loci involved in maize evolution. *Genet Res* 74: 291–302.

505 MALOOF, M., K. SOLIMAN, R. JORGENSEN, AND R. ALLARD. 1984. Ribosomal dna spacer
 506 length polymorphisms in barley - mendelian inheritance, chromosomal location, and population
 507 dynamics. *P Natl Acad Sci Usa* 81: 8014–8018.

508 MATSUOKA, Y., Y. VIGOUROUX, M. GOODMAN, G. SANCHEZ, E. BUCKLER, AND
 509 J. DOEBLEY. 2002. A single domestication for maize shown by multilocus microsatellite
 510 genotyping. *P Natl Acad Sci Usa* 99: 6080–6084.

511 MOELLER, D. A., M. I. TENAILLON, AND P. TIFFIN. 2007. Population structure and its effects
 512 on patterns of nucleotide polymorphism in teosinte (*zea mays* ssp. *parviglumis*). *Genetics* 176:
 513 1799–1809.

514 OLSEN, K. AND B. GROSS. 2010. Genetic perspectives on crop domestication. *Trends in Plant*
515 *Science* 15: 529–537.

516 OLSEN, K. M. AND J. F. WENDEL. 2013. A bountiful harvest: Genomic insights into crop
517 domestication phenotypes. *Annual Review of Plant Biology* 64: 47–70.

518 PIPERNO, D., A. RANERE, I. HOLST, J. IRIARTE, AND R. DICKAU. 2009. Starch grain and
519 phytolith evidence for early ninth millennium bp maize from the central balsas river valley,
520 mexico. *P Natl Acad Sci Usa* 106: 5019–5024.

521 PLANTINGA, T., S. ALONSO, N. IZAGIRRE, M. HERVELLA, R. FREGEL, J. VAN DER MEER,
522 M. NETEA, AND C. DE LA RUA. 2012. Low prevalence of lactase persistence in neolithic
523 south-west europe. *Eur J Hum Genet* 20: 778–782.

524 PYHÄJÄRVI, T., M. HUFFORD, AND J. ROSS-IBARRA. 2013. Complex patterns of local
525 adaptation in the wild relatives of maize. *Genome Biology and Evolution* 5: 1594–1609.

526 ROSS-IBARRA, J., P. MORRELL, AND B. GAUT. 2007. Plant domestication, a unique
527 opportunity to identify the genetic basis of adaptation. *P Natl Acad Sci Usa* 104: 8641–8648.

528 ROSS-IBARRA, J., M. TENAILLON, AND B. GAUT. 2009. Historical divergence and gene flow in
529 the genus *zea*. *Genetics* 181: 1399–1413.

530 ROZEN, S. AND H. SKALETISKY. 2000. Primer3 on the www for general users and for biologist
531 programmers. *Methods in Molecular Biology* : 365–386.

532 SIGMON, B. AND E. VOLLBRECHT. 2010. Evidence of selection at the *ramosa1* locus during
533 maize domestication. *Mol Ecol* 19: 1296–1311.

534 STUDER, A. AND J. DOEBLEY. 2012. Evidence for a natural allelic series at the maize
535 domestication locus *teosinte branched1*. *Genetics* 19: 951–958.

536 STUDER, A., Q. ZHAO, J. ROSS-IBARRA, AND J. DOEBLEY. 2011. Identification of a functional
537 transposon insertion in the maize domestication gene *tb1*. *Nat Genet* 43: 1160–U164.

538 THORNTON, K. 2003. libsequence: a c++ class library for evolutionary genetic analysis.
539 *Bioinformatics* 19: 2325–2327.

540 THURBER, C., M. REAGON, B. GROSS, K. OLSEN, Y. JIA, AND A. CAICEDO. 2010. Molecular
 541 evolution of shattering loci in us weedy rice. *Mol Ecol* 19: 3271–3284.

542 TISHKOFF, S., F. REED, A. RANCIARO, B. VOIGHT, C. BABBITT, J. SILVERMAN, K. POWELL,
 543 H. MORTENSEN, J. HIRBO, M. OSMAN, M. IBRAHIM, S. OMAR, G. LEMA, T. NYAMBO,
 544 J. GHORI, S. BUMPSTEAD, J. PRITCHARD, G. WRAY, AND P. DELOUKAS. 2007. Convergent
 545 adaptation of human lactase persistence in africa and europe. *Nat Genet* 39: 31–40.

546 VAN HEERWAARDEN, J., J. DOEBLEY, W. BRIGGS, J. GLAUBITZ, M. GOODMAN,
 547 J. GONZALEZ, AND J. ROSS-IBARRA. 2011. Genetic signals of origin, spread, and introgression
 548 in a large sample of maize landraces. *P Natl Acad Sci Usa* 108: 1088–1092.

549 WARBURTON, M. L., W. GARRISON, S. TABA, A. CHARCOSSET, C. MIR, F. DUMAS,
 550 D. MADUR, S. DREISIGACKER, C. BEDOYA, B. PRASANNA, C. XIE, S. HEARNE, AND
 551 J. FRANCO. 2011. Gene flow among different teosinte taxa and into the domesticated maize
 552 gene pool. *Genetic Resources and Crop Evolution* 58: 1243–1261.

553 WEBER, A., R. CLARK, L. VAUGHN, J. SANCHEZ-GONZALEZ, J. YU, B. YANDELL,
 554 P. BRADBURY, AND J. DOEBLEY. 2007. Major regulatory genes in maize contribute to
 555 standing variation in teosinte (*zea mays* ssp *parviglumis*). *Genetics* 177: 2349–2359.

556 WHIPPLE, C., T. KEBROM, A. WEBER, F. YANG, D. HALL, R. MEELEY, R. SCHMIDT,
 557 J. DOEBLEY, T. BRUTNELL, AND D. JACKSON. 2011. grassy tillers1 promotes apical
 558 dominance in maize and responds to shade signals in the grasses. *P Natl Acad Sci Usa* 108:
 559 E506–E512.

560 WILKES, H. 1977. Hybridization of maize and teosinte, in mexico and guatemala and the
 561 improvement of maize. *Economic Botany* 31: 254–293.

562 WRIGHT, S. I., I. V. BI, S. G. SCHROEDER, M. YAMASAKI, J. F. DOEBLEY, M. D.
 563 McMULLEN, AND B. S. GAUT. 2005. The effects of artificial selection on the maize genome.
 564 *Science* 308: 1310–1314.

565 ZHANG, L., Q. ZHU, Z. WU, J. ROSS-IBARRA, B. GAUT, S. GE, AND T. SANG. 2009. Selection
 566 on grain shattering genes and rates of rice domestication. *New Phytol* 184: 708–720.

Table 1. Pairwise F_{ST} values from sequence and *Hopscotch* genotyping data

Comparison	Region 1	Region 2	<i>Hopscotch</i>
EjuA & EjuB	0	0	0
EjuA & MSA	0.326	0.328	0.186
EjuA & SLO	0.416	0.258	0.280
EjuB & MSA	0.397	0.365	0.188
EjuB & SLO	0.512	0.290	0.280
MSA & SLO	0.007	0	0.016

Table 2. Population genetic statistics from resequenced regions near the *tb1* locus

Population	# Haplotypes	Hap. Diversity	$\hat{\theta}_\pi$	Tajima's D
<i>Region 1(5' UTR)</i>				
EJUA	8	0.859	0.005	-1.650
EJUB	5	0.709	0.004	-1.831
MSA	6	0.682	0.004	-1.755
SLO	3	0.318	0.001	-0.729
<i>Region 2 (66kb upstream)</i>				
EJUA	8	0.894	0.018	0.623
EJUB	8	0.894	0.016	0.295
MSA	3	0.682	0.011	-0.222
SLO	4	0.742	0.014	0.932

Table 3. mean r^2 values between SNPs on chromosome 1, in the broad *tb1* region, within the 5' UTR of *tb1* (Region 1), and 66kb upstream of *tb1* (Region 2).

Population	Chr. 1	<i>tb1</i> region	Region 1	Region 2
Ejutla A	0.095	0.050	0.747	0.215
Ejutla B	0.069	0.051	0.660	0.186
La Mesa	0.070	0.053	0.914	0.766
San Lorenzo	0.101	0.067	0.912	0.636

Table 4. Assignments to maize and teosinte in the *tb1* and chromosome 1 regions from STRUCTURE

Population	<i>tb1</i> region		Chr 1	
	Maize	Teosinte	Maize	Teosinte
Ejutla A	0.022	0.978	0.203	0.797
Ejutla B	0.019	0.981	0.187	0.813
La Mesa	0.012	0.988	0.193	0.807
San Lorenzo	0.016	0.984	0.205	0.795

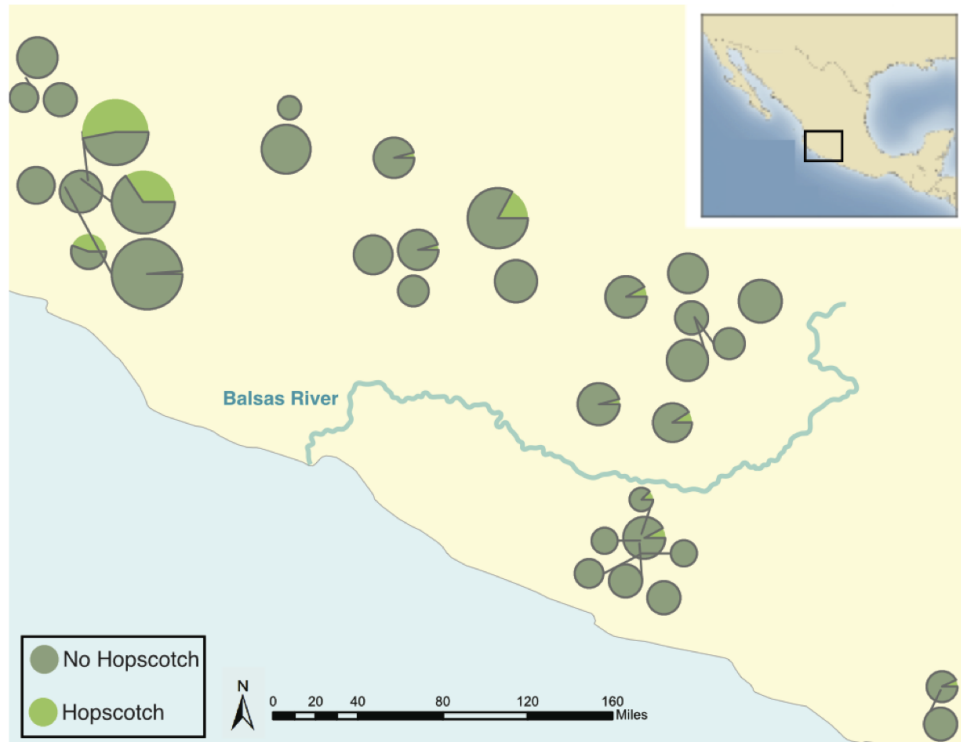


Figure 1. Map showing the frequency of the *Hopscotch* allele in populations of *parviglumis* where we sampled more than 6 individuals. Size of circles reflects number of individuals sampled. The Balsas River is shown, as the Balsas River Basin is believed to be the center of domestication of maize.

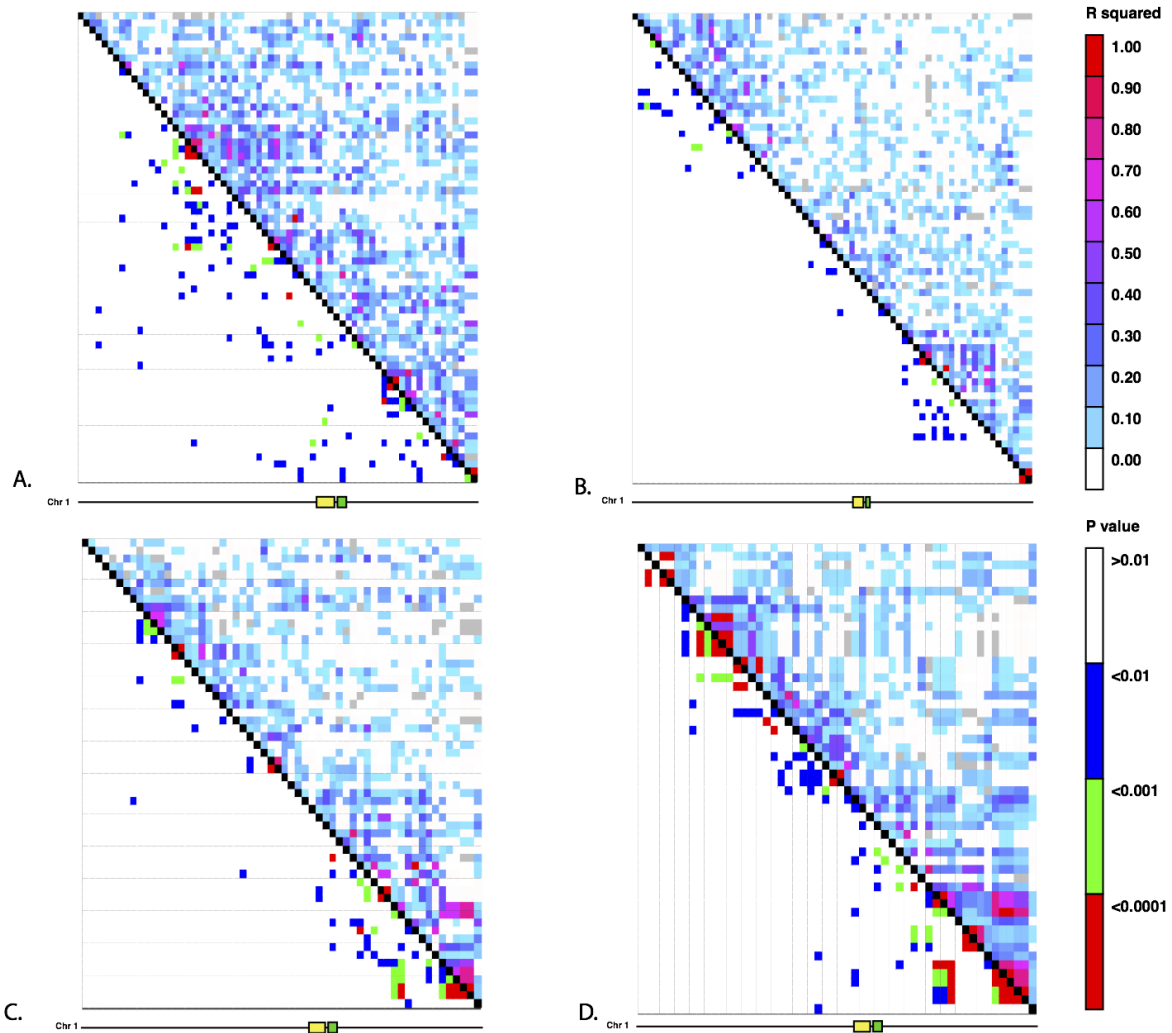


Figure 2. Linkage disequilibrium for SNPs in Mb 261-268 on chromosome 1. The yellow rectangle indicates the location of the *Hopscotch* insertion and the green rectangle represents the *tb1* ORF. A) Ejutla A; B) Ejutla B; C) La Mesa; D) San Lorenzo. The upper triangle above the black diagonal is colored based on the r^2 value between SNPs while the bottom triangle is colored based on p-value for the corresponding r^2 value.

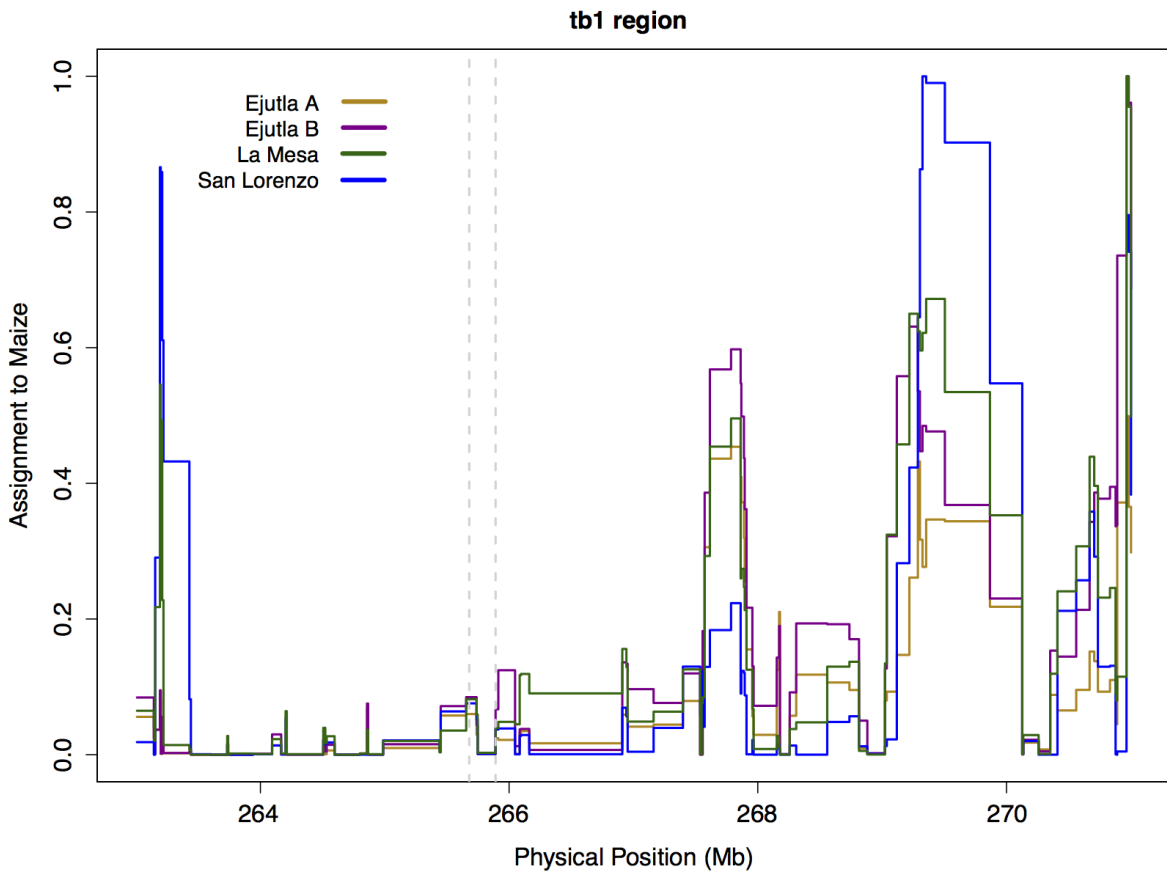


Figure 3. STRUCTURE assignment to maize across a section of chromosome 1. The dotted lines mark the beginning of the sequenced region 66kb upstream (Region 2) and the end of the *tb1* ORF.

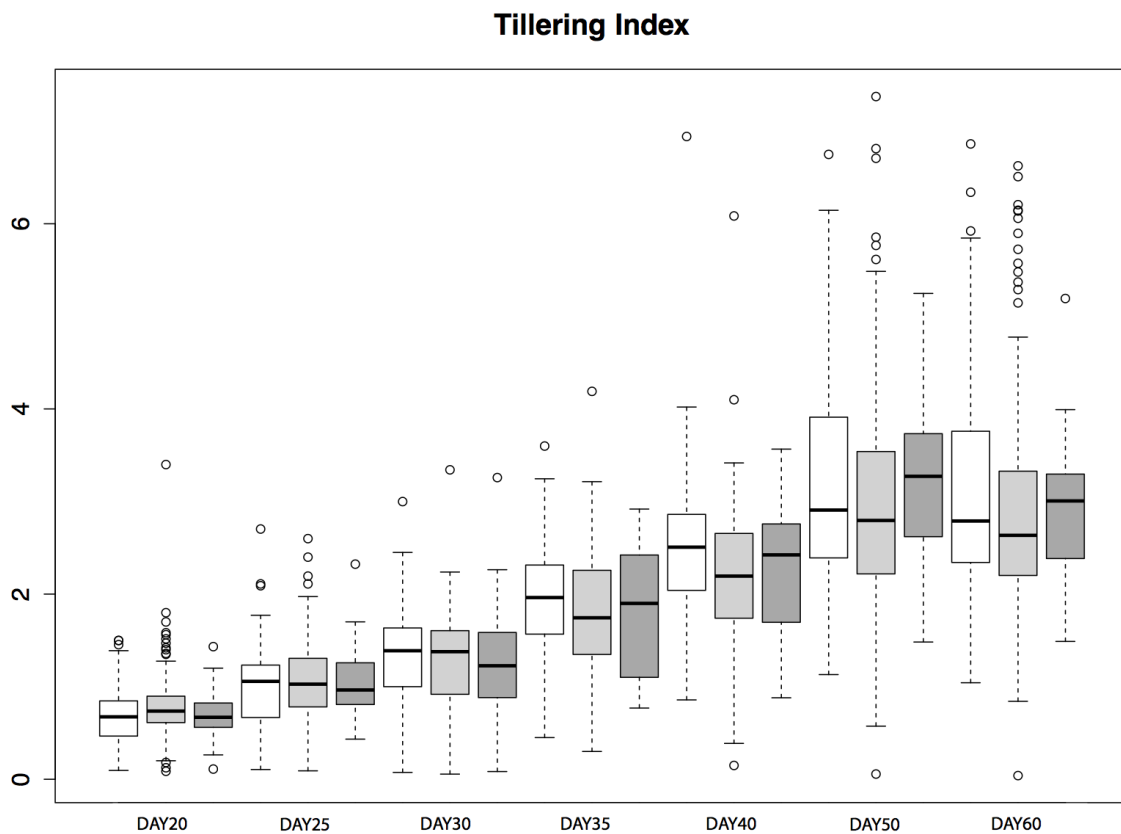


Figure 4. Box-plots showing tillering index in greenhouse grow-outs for phenotyping. White indicates individuals homozygous for the *Hopscotch*, light grey represents heterozygotes, and dark grey represents homozygotes for the teosinte (No *Hopscotch*) allele. Within boxes, dark black lines represent the median, and the edges of the boxes are the first and third quartiles. Outliers are displayed as dots, the maximum value excluding outliers is shown with the top whisker, while the minimum excluding outliers is shown with the bottom whisker.