Natural variation in teosinte at the domestication locus teosinte branched1 (tb1)

Laura Vann[1], Thomas Kono[1,2], Tanja Pyhäjärvi[1,3], Matthew B. Hufford[1,4], and Jeffrey Ross-Ibarra*[1,5]

[1]Department of Plant Sciences, University of California Davis

[2]Department of Agronomy and Plant Genetics, University of Minnesota Twin Cities

[3]Department of Biology, University of Oulu

[4]Department of Ecology, Evolution, and Organismal Biology, Iowa State University

[5]Center for Population Biology and Genome Center, University of California Davis

April 6, 2014

*rossibarra@ucdavis.edu

Abstract

The *teosinte branched1* (*tb1*) gene, a repressor of lateral organ growth, is a major QTL involved in branching differences between maize and its wild progenitor, teosinte. Further studies have shown that the insertion of a transposable element (Hopscotch) upstream of *tb1* enhances its expression, causing the reduction in branching observed in domesticated maize. Observations of the maize *tb1* allele in teosinte individuals, coupled with estimates of the age of insertion of the Hopscotch element, led us to investigate the role of *tb1* in teosinte. Results from genotyping across many natural populations suggest that the Hopscotch element is segregating at a higher than expected frequency in a number of populations of both subspecies *parviglumis* and subspecies *mexicana*. Analysis of linkage disequilibrium between the Hopscotch element and variation (SNPs) in surrounding regions does not support a hypothesis of recent introgression from maize into teosinte, and we find no evidence of environmental correlations that might suggest recent selection. Finally, two greenhouse experiments fail to find an important role for *tb1* in controlling tillering in natural populations of *parviglumis*. Our findings suggest that the role of the Hopscotch in tillering in teosinte is not as straightforward as it is in domesticated maize, and other loci known to affect branching architecture in maize may play a more important role in tillering in teosinte.

Introduction

Domesticated crops and their wild progenitors provide an excellent system in which to study adaptation and genomic changes associated with human-mediated selection (41). Perhaps the central focus of the study of domestication has been the identification of genetic variation underlying agronomically important traits such as fruit size and plant architecture (37). Additionally, many domesticates show reduced genetic diversity when compared to their wild progenitors, and an understanding of the distribution of diversity in the wild and its phenotypic effects has become increasingly useful to crop improvement (**?**). But while some effort has been invested into understanding how wild alleles behave in their domesticated relatives, very little is known about the role that alleles found most commonly in domesticates play in natural populations of their wild progenitors.

Maize (*Zea mays* ssp. *mays*) was domesticated from the teosinte *Zea mays* ssp. *parviglumis* (hereafter, *parviglumis*) roughly 10,000 B.P. in southwest Mexico (38; 35). Domesticated 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 (27; 10; 29). Additionally, large naturally occurring populations of both *Zea mays* ssp. *parviglumis* (the wild progenitor of maize) and *Zea mays* ssp. *mexicana* (highland teosinte) can be found throughout Mexico (53; 28), and genetic diversity and effective population size (Ne) of these taxa is estimated to be high (42) relative to other plants (36).

Many morphological changes are associated with the domestication of maize, and understanding the genetic basis for these changes has been a focus of maize research for a number of years. 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 differences between maize and teosinte (14; 11; 33). Further studies have shown that the insertion of a 4.9 kb retrotransposon (Hopscotch) in the upstream control region of *tb1* led to increased expression of this gene, causing the reduction in branching observed in domesticated maize (45). The effects of this insertion have been observed in tiller number in maize, but little is known about its role, if any, in teosinte (45). Dating of this element has suggested that its insertion predates the domestication of maize, leading to the hypothesis that it was segregating as standing variation in ancient populations of teosinte and increased to high frequency in maize due to selection during domestication (45). Furthermore, (44) investigated the phenotypic effects of 9 teosinte *tb1* alleles in an isogenic maize background and found that the introgressions sort into three distinct phenotypic classes, suggesting that variation at the *tb1* locus may play a functional role in teosinte. Tillering, or more specifically lack of tillers in teosinte, may provide an ecological advantage to plants. In species such as teosinte, where plants grow densely packed together, if plants invest more resources into growing taller and fewer resources into tillering, plants may be able to out-compete neighbors for limiting resources such as light. In natural populations of teosinte variation at the (tb1) locus may then play a role in the ecology of teosinte. In this study we aim to characterize the distribution of the Hopscotch insertion in *mexicana*, *parviglumis*, and landrace maize, and to examine the phenotypic effects of the insertion in *parviglumis*. We use a combination of PCR genotyping for the Hopscotch element in our full panel and sequencing of two small regions upstream of *tb1* in a subset of teosinte populations to explore patterns of genetic variation at this locus. Finally, we test for an association between the

Hopscotch element and tillering phenotypes in a population of *parviglumis*.

Methods

Sampling and Genotyping

We sampled 1,110 individuals from 350 accessions (247 maize landraces, 17 *Zea mays* ssp. *mexicana* (hereafter *mexicana*) populations, and 86 *parviglumis* populations) and assessed the presence or absence of the Hopscotch insertion (**??**, **??**). DNA was extracted from leaf tissue using a modified CTAB approach (15; 34). We designed primers using PRIMER3 (43) implemented in Geneious (30) to amplify the entire Hopscotch element, as well as an internal primer allowing us to simultaneously check for possible PCR bias between presence (5 kb amplification product) and absence (300 bp amplification product) of the Hopscotch insertion. Two PCRs were performed for each individual, one with primers flanking the Hopscotch (HopF/HopR) and one with a flanking primer and an internal primer (HopF/HopIntR). Primer sequences are HopF, 5'-TCGTTGATGCTTTGATGGATGG-3'; Hop R, 5'-AACAGTATGATTTCATGG
3'; and HopIntR, 5'-CCTCCACCTCTCATGAGATCC-3' (1). Homozygotes show a single band for either the Hopscotch element (5 kb) or the absence of the element (300 bp), while heterozygotes are three-banded, showing a band for both the presence and absence of the Hopscotch element as well as a band for the internal primer set (2). When only one PCR resolved well, we scored one allele for the individual, which explains the odd number of alleles included in our analyses. We used Phusion High Fidelity Enzyme (Finnzymes, Inc.) and the following conditions for amplifications: 98C for 3 min, 30 cycles of 98C for 15 s, 65C for 30 s, and 72C for 3 min 30 s, with a final extension of 72C for 10 min. PCR products were visualized on a 1%

agarose gel and scored for presence/absence of the Hopscotch based on band size.

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* (sequenced region 1) and from 1,235 bp upstream of the start of the Hopscotch and 66,169 bp upstream from the start of the *tb1* ORF (sequenced region 2). The 5' UTR (containing sequenced region 1) has been shown to have elevated diversity in ssp. *parviglumis* with respect to maize, while the the area upstream from the Hopscotch (sequenced region 2) has been shown to be critical in determining basal branching and ear architecture (50; 6). We designed the following primers using PRIMER3 (43): for the 5' UTR, 5' GGATAATGTGCACCAGGTGT 3' and 5' GCGTGCTAGAGACACYTGTTGCT 3'; for the 50 kb upstream region, 5' TGTCCTCGCCGCAACTC 3' and 5' TGTACGCCCGCCCCTCATCA 3' (1). We used Taq Polymerase (New England Biolabs) and the following thermal cycler conditions to amplify fragments: 94C for 3 min, 30 cycles of 92C for 40 s, annealing for 1 min, 72C for 40 s, and a final 10 min extension at 72C. Annealing temperatures for sequenced region 1 and sequenced region 2 were 59.7C and 58.8C, respectively. To clean excess primer and dNTPs we added two units of Exonuclease1 and 2.5 units of Antarctic Phosphatase to 8.0 L of amplification product. This mix was placed on a thermal cycler with the following program: 37C for 30 min, 80C for 15 min, and a final cool-down step to 4C.

We cloned cleaned fragments into a TOPO-TA vector (Invitrogen, Carlsbad)

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: 96C for 5 min; then 35 cycles at 96C for 30 s, 53C for 30 s, 72C for two min; and a final extension at 72C 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 (30). 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.

Genotyping Analysis

We examined discrepancies between observed and expected genotype frequencies by calculating Hardy-Weinberg Equilibrium (HWE), and to calculate differentiation between populations (F_{ST}) and subspecies (F_{CT}) we used HierFstat (24). These analyses only included populations in which 8 or more individuals were sampled. To test the hypothesis that the Hopscotch insertion may be adaptive under certain

environmental conditions, we looked for significant associations between the Hopscotch frequency and environmental variables using BayEnv (9). 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 (Bayes Factor)(9). We used genotyping and covariance data from Pyhäjärvi et al., (2013) (40) for BayEnv, with the Hopscotch insertion coded as an additional SNP (3). Environmental data were obtained from www.worldclim.org, the Harmonized World Soil Database and www.harvestchoice.org, and summarized by principle component analysis Pyhjrvi et al. (40) (40).

Sequence Analysis

For population genetic analyses of sequenced region 1 and sequenced region 2 we used the analysis package of Libsequence (46) to calculate pairwise F_{ST} between populations, and to calculate standard diversity statistics (number of haplotypes; haplotype diversity; Watterson's estimator θ_W; pairwise nucleotide diversity θ_π; and Tajima's D). To produce a visual representation of differentiation between sequences and to examine patterns in sequence clustering by Hopscotch genotype we used Phylip (http://evolution.genetics.washington.edu/phylip.html) to create neighbor-joining trees with bootstrapping (100 repetitions) to examine the support of nodes in our trees. For creation of trees we also included homologous sequence data from 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 and landraces (data from Maize HapMapV2, (5)).

In order to assess patterns of linkage disequilibrium (LD) around the Hopscotch element in the context of chromosomal patterns of LD we used Tassel (3) and calculated LD between SNPs across chromosome 1 using previously published data from twelve plants each of the Ejutla A, Ejutla B, San Lorenzo, and La Mesa populations (40). 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 San Lorenzo, Ejutla B, Ejutla A, and La Mesa, respectively. We then used Tassel (3) 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 (18) and the same phased 55K SNP data from (40) that we used for LD analysis, combined with the corresponding SNP data from a diverse panel of 282 maize lines (8). SNPs were anchored in a modified version of the IBM genetic map ((23), `http://arxiv.org/abs/1307.7313`). We created haplotype blocks using a custom Perl script that grouped SNPs separated by less than 5kb into haplotypes. We ran STRUCTURE at K=2 under the linkage model, performing 3 replicates with an MCMC burn-in of 10,000 steps and 50,000 steps post burn-in.

Phenotyping of *Zea mays.* ssp. *parviglumis*

To investigate the phenotypic effects of the Hopscotch insertion in teosinte, we conducted an initial phenotyping trial (Phenotyping 1). We germinated 250 seeds of *parviglumis* collected in Jalisco state, Mexico (population San Lorenzo) (26) where

the Hopscotch is segregating at highest frequency (0.4375) 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 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 fungicide and germinated them in petri dishes with filter paper. Successful germinations (206 individuals) were then planted into one gallon size pots with potting soil and randomly spaced one foot apart on greenhouse benches. Plants were watered three times a day with an automatic drip containing 10-20-10 fertilizer.

Starting on day 15, we measured tillering index, the ratio of the sum of tiller lengths to the height of the plant (4). Tillering index has been shown to be the most effective way to observe the phenotypic effects of the Hopscotch insertion on plant architecture in maize (4). Following initial measurements, we phenotyped plants for tillering index every 5 days through day 40, and then on day 50 and day 60. On day 65 we measured culm diameter between the third and fourth nodes of each plant. Culm diameter is not believed to be correlated with tillering index, or variation at *tb1* (e.g. Hopscotch genotype). Following phenotyping we extracted DNA from all plants using a modified SDS extraction protocol (http://www.ars.usda.gov). We genotyped individuals for the Hopscotch insertion following the protocols listed above. Based on these initial data, we conducted a post hoc power analysis using data from day 40 of phenotyping 1, indicating that a minimum of 71 individuals in each genotype class are needed to detect the observed effect of the Hopscotch on tillering index.

We performed a second phenotyping experiment (phenotyping 2) in which we germinated 372 seeds of *parviglumis*, choosing equally between sites previously determined to have or not have the Hopscotch insertion. Seeds were germinated and planted on day 7 post fruit-case removal into 2 gallon pots. Plants were watered

twice daily, alternating between fertilized and non-fertilized water. We began phenotyping successful germinations (302) for tillering index on day 15 post fruit case removal, and phenotyped every five days until day 50. At day 50 we measured culm diameter between the third and fourth nodes. We extracted DNA and genotyped plants following the same guidelines as in phenotyping 1.

Resulting tillering index data for each genotype class did not meet the criteria for a repeated measures ANOVA, so we transformed the data using a Box-Cox transformation ($\alpha = 0$) implemented in the car package in R (20) to improve the normality and homogeneity of variance among genotype classes. We analyzed relationships between genotype and tillering index and tiller number using a repeated measures ANOVA through a general linear model function implemented in SAS v.9.3 (SAS Institute Inc., Cary, NC, USA). Additionally, in order to compare any association between Hopscotch genotype and tillering and associations at other presumably unrelated traits, we performed an ANOVA between culm diameter and genotype using the same general linear model in SAS.

Results

Genotyping

We genotyped 247 accessions of maize landraces, and found the Hopscotch element fixed in all but 8 of these accessions (**??**, **??**). For all analyses we only included individuals that clearly resolved for both PCRs, resulting in 837 individuals total. Within our *parviglumis* and *mexicana* samples we found the Hopscotch insertion segregating in 37 and 4 populations, respectively, and at highest frequency in the states of Jalisco, Colima, and Michoacán in central-western Mexico in both subspecies (1).

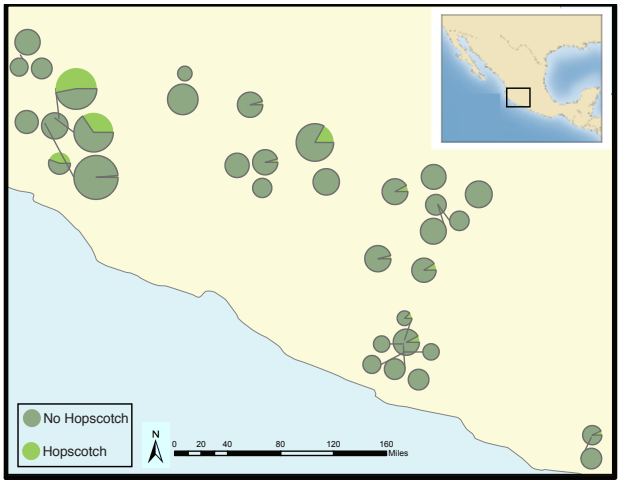


Figure 1: Map showing the frequency of the Hopscotch allele in populations of *Zea mays* ssp. *parviglumis* where we sampled more than 6 individuals. Size of circles reflects number of alleles sampled.

We examined Hardy-Weinberg equilibrium in a total of 14 populations (10 *parviglumis* and 4 *mexicana*) with more than 8 individuals sampled per population. Three populations (RIMPA0073, RIMPA0093, and RIMPA0158) show evidence of deviations from expected genotype frequencies under the assumptions of HWE (p¡0.05).

Using our Hopscotch genotyping data, we calculated differentiation between pop-

ulations (F_{ST}) and subspecies (F_{CT}) for populations in which we sampled 8 or more alleles. F_{CT} is 0 within our dataset, and we found similar levels of F_{ST} between populations within each subspecies (0.22) and between all populations (0.23), to those reported in genome-wide estimates from previous studies (40) (1). Although

Table 1: Pairwise F_{CT} values from sequence data

Comparison	**Sequenced Region 2**	**Sequenced Region 1**	**Hopscotch**
EjuA & EJuB	0	0	0
EjuA & MSA	0.328	0.326	0.186
EjuA & SLO	0.258	0.416	0.28
EjuB & MSA	0.365	0.397	0.188
EjuB & SLO	0.29	0.512	0.28
MSA & SLO	0	0.007	0.016

we found large variation in Hopscotch allele frequency among our populations, our BayEnv analysis did not indicate a correlation between the Hopscotch insertion and environmental variables (all Bayes Factors < 1; 3).

Sequencing

To investigate patterns of sequence diversity and linkage disequilibrium (LD) in the *tb1* region, we sequenced two small (<1kb) regions upstream of the tb1 ORF in four populations. After alignment and singleton checking we recovered 40 and 48 segregating sites for the 50kb upstream region and the 5' UTR region, respectively. For region 1, Ejutla A has the highest values of haplotype diversity, θ_W, 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 similarly negative in the

two Ejutla populations and La Mesa, but is more positive in San Lorenzo (2).

Table 2: Add caption

Population	Number of Haplotypes	Haplotype Diversity	θ_W	θ_π	Tajima's D
		Sequenced region 2 (50kb upstream)			
EJUA	8	0.89394	0.01548	0.01763	0.6231
EJUB	8	0.89394	0.01493	0.01591	0.29504
MSA	3	0.68182	0.01111	0.01055	-0.22212
SLO	4	0.74242	0.01167	0.01413	0.93185
		Sequenced region 1(5? UTR)			
EJUA	8	0.85897	0.00874	0.00527	-1.64955
EJUB	5	0.70909	0.00663	0.00378	-1.83123
MSA	6	0.68182	0.00646	0.00373	-1.75506
SLO	3	0.31818	0.00176	0.00137	-0.72873

For region 2, haplotype diversity, θ_W, and θ_π, are similar for Ejutla A and Ejutla B, while La Mesa and San Lorenzo hav

The teosinte populations with the highest frequency of the Hopscotch insertion in this study were 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 one, and specifically in the *tb1* region. If the Hopscotch is found in these populations due to recent introgression 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 (2), and r^2 in the *tb1* region (positions 264,596,664-265,891,456; AGPv2) does not differ significantly

between populations with (average r^2 of 0.085) and without the Hopscotch genotype (average $r^2 = 0.082$). In fact, average r^2 is lower in the *tb1* region ($r^2 = 0.056$) than across the rest of chromosome 1 ($r^2 = 0.083$) (3).

Table 3: r^2 values between SNPs in the *tb1* region (positions 264,596,664-265,891,456) and the rest of chromosome 1, within the 5? UTR (Sequenced region 1), and within the 66,169 bp upstream region (Sequenced region 2).

2[2]*Population	2[2]*Chromosome 1	2[2]*tb1 region	2[2]*Sequenced region 1	2[2]*Sequenced re
Ejutla A	0.095426101	0.050304	0.747295	0.214933
Ejutla B	0.068681837	0.051295	0.660354	0.186395
La Mesa	0.069500533	0.053306	0.914286	0.766234
San Lorenzo	0.100536784	0.067251	0.912281	0.636364

The lack of clustering of Hopscotch genotypes in our NJ tree as well as the lack of LD around *tb1* does not support the hypothesis that the Hopscotch insertion in these populations of *parviglumis* is the result of recent introgression. However, to further explore this hypothesis we performed a STRUCTURE analysis using Illumina MaizeSNP50 data from four of our *parviglumis* populations (EjuA, EjuB, MSA, and SLO) and the maize 282 diversity panel (8;). The linkage model implemented in STRUCTURE can be used to identify ancestry of blocks of linked variants, which would arise as a result of recent admixture between populations. If the Hopscotch insertion is present in populations of *parviglumis* as a result of recent admixture with domesticated maize, we would expect the insertion and linked variants in surrounding sites to be assigned to the "maize" cluster in our STRUCTURE runs, not the "teosinte" cluster. In all runs, assignment to maize in the *tb1* region across all four

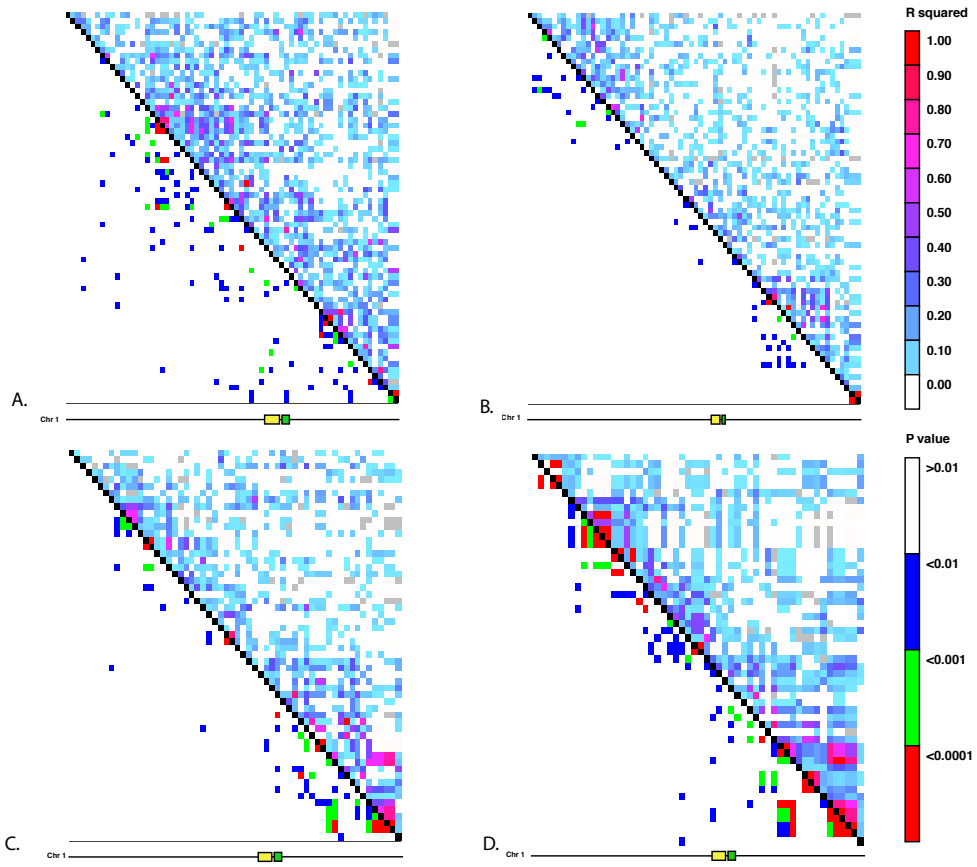


Figure 2: Linkage equilibrium plots for SNPs within bp 261,000,000-268,000,000 on chromosome 1. The yellow rectangle indicates the location of the Hopscotch insertion and the green represents the *tb1* ORF. A) Ejutla A; B) Ejutla B; C) La Mesa; D). San Lorenzo

parviglumis populations is lower (average 0.017) than the chromosome wide average (0.20; 3).

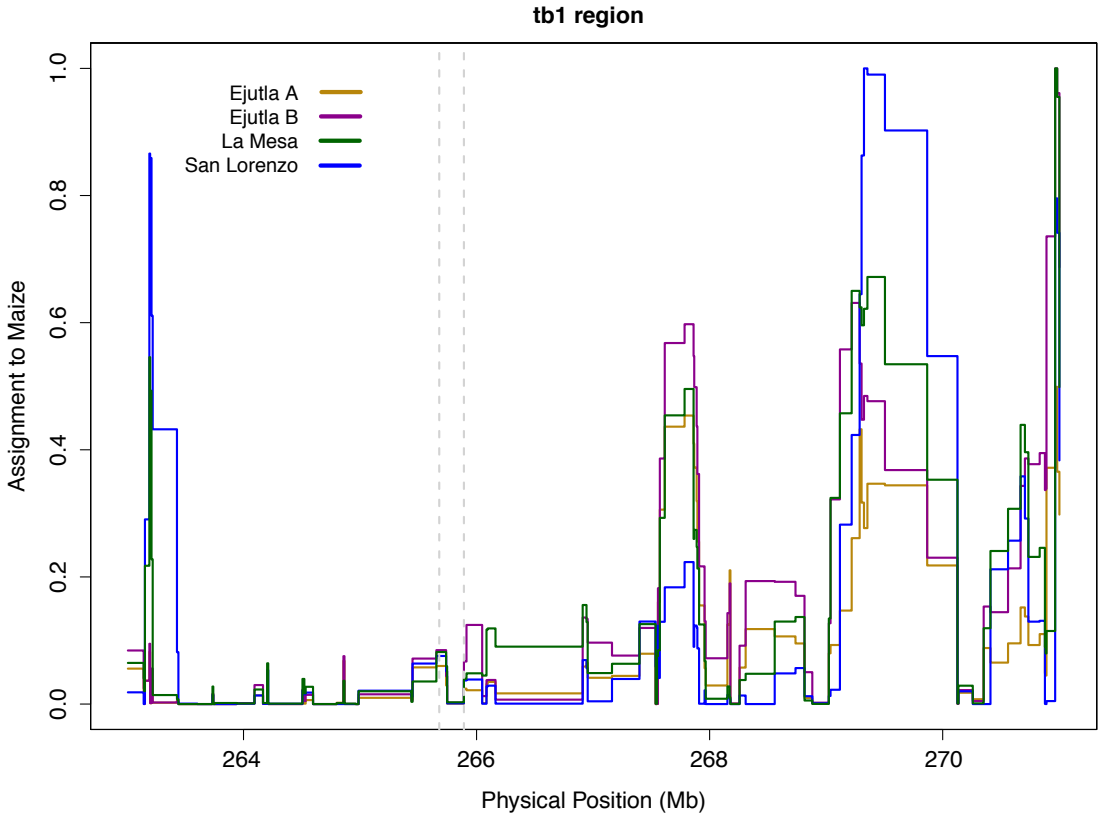


Figure 3: STRUCTURE assignment to maize across a section of chromosome 1. The dotted lines mark the beginning of the sequenced region 50kb upstream (Sequenced region 2) and the end of the *tb1* ORF.

Phenotyping

Phenotyping 1

We measured tillering index (TI), the ratio of the sum of tiller lengths to plant height, for 216 plants from 23 sampling locations within the San Lorenzo population, and genotyped plants for the Hopscotch insertion. We found the Hopscotch segregating

at a frequency of 0.65; with a 0.4 frequency of homozygotes for the Hopscotch, a 0.5 frequency of heterozygotes, and a 0.1 frequency of homozygotes for the teosinte allele (no Hopscotch), showing no significant deviations from expected frequencies under Hardy-Weinberg equilibrium. After performing a repeated measures ANOVA between our transformed tillering index data and Hopscotch genotype we find a weak positive correlation between presence of the Hopscotch and tillering index on day 40 (p=0.0848), but no correlation between tillering index and genotype on any other day (4). Additionally we find no significant correlation between tiller number and Hopscotch genotype, or culm diameter and Hopscotch genotype in phenotyping 1.

Phenotyping 2

For phenotyping 2 we measured tillering index every five days through day 50 for 302 plants. We followed the same transformations and analyses as described in phenotyping 1. We find the Hopscotch allele segregating at a frequency of 0.69, with a 0.6 frequency of Hopscotch homozygotes, and a 0.2 frequency of both heterozygotes and homozygotes for the teosinte allele. We find similar patterns as in phenotyping 1, with a weak positive correlation between tillering index and Hopscotch genotype at day 40 (p¡0.0611), with no significant correlation on any day. Similarly, relationships between Hopscotch genotype and tiller number, and hopscotch genotype and culm diameter are not significant.

Discussion

Adaptation occurs either due to selection on standing variation or on *de novo* mutations. Adaptation as a result of selection on standing variation has been well-described in a number of systems, for example, selection for lactose tolerance in

humans (39; 48); variation at the *Eda* locus in three-spined stickleback (31; 7); and pupal diapause in the Apple Maggot fly (19). Although the role of standing variation with respect to adaptation has been described in many systems, its importance to domestication is not as well studied.

In maize, alleles at important domestication loci (*RAMOSA1*, (**?**); *barren stalk1*, (22); and *grassy tillers1*, (52)) have been shown to have been selected from standing variation, suggesting that diversity already present in teosinte may have played an important role in the domestication of maize. The *teosinte branched1* gene has long been a central focus of research concerning maize domestication, and, while previous studies have suggested that differences in plant architecture between domesticated maize and teosinte are a result of selection on standing variation, little is known about variation at this locus in teosinte (6; 45). Studer et al. (45) genotyped 90 accessions of teosinte (inbred and outbred), and provided the first evidence that the Hopscotch insertion is segregating in teosinte (45).

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, in sampling numerous individuals from many teosinte populations our study provides greater insight into the distribution and prevalence of the Hopscotch in teosinte. While our findings are consistent with a previous study by Studer et al. (45) in that we identified the Hopscotch allele segregating in teosinte, we find it at higher frequency than previously suggested (45). Crop to wild introgression has previously been observed at domestication loci ((17; 54; 47; 1; 25; 53; 49; 2)), but our results are more consistent with Hufford et al. (28) who found resistance to introgression from maize into teosinte (28). Furthermore, Hufford et al. (28) showed that domestication loci, such as *tb1*, are particularly resistant to introgression in both directions of gene flow (i.e., maize to teosinte and teosinte to maize) (28).

We find no evidence of recent introgression in our analyses. Clustering patterns in our NJ trees do not reflect a pattern expected if maize alleles at the *tb1* locus had introgressed into populations of teosinte. We do not have sympatric maize samples from many of our populations, however, so we cannot rule out the possibility that Hopscotch alleles in these teosinte populations are segregating in sympatric maize. Moreover, analysis of linkage in the *tb1* region does not reveal patterns of high LD relative to the rest of chromosome 1. The similarity in LD in the *tb1* region in populations with and without the Hopscotch provides further support for an explanation other than introgression. Additionally, assignment to maize in this region in our STRUCTURE analysis is in fact lower than the average across chromosome 1 (3, 4).

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

	tb1 region		Chr 1	
Population	Maize assignment	Teosinte assignment	Maize assignment	Teosinte assignment
Ejutla A	0.02158681	0.9784132	0.2026814	0.7973186
Ejutla B	0.01888194	0.9811181	0.1872131	0.8127869
La Mesa	0.9881333	0.0118675	0.8068998	0.1931017
San Lorenzo	0.01551389	0.9844861	0.2048252	0.7951748

Although recent introgression seems unlikely, we cannot rule out ancient introgression as an explanation for the presence of the Hopscotch in these populations. In the ancient introgression scenario, recombination would have broken up linked variants surrounding the Hopscotch and we would find LD patterns similar to the ones we report here. However, it seems unlikely that there would have been ancient introgression between these populations of teosinte and their sympatric maize pop-

ulations and not present-day introgression. Studies have demonstrated the presence of ongoing gene flow between domesticated maize and both *Zea mays* ssp. *mexicana* and *Zea mays* ssp. *parviglumis* in a number of sympatric populations (28; 16**?**).

Deviations from expected genotype frequencies under Hardy-Weinberg equilibrium at the Hopscotch insertion in the San Lorenzo population and additional populations may be indicative of demographic changes that have affected genome-wide patterns of variation, or may be the result of selective pressures acting on or near the *tb1* locus. Previous studies using both SSRs and genome-wide SNP data have provided evidence for a population bottleneck in the San Lorenzo population that has resulted in lower levels of diversity across the genome (26;). The lower levels of sequence diversity in the 5' UTR region and the more positive values of Tajima's D in the San Lorenzo population that we present here are consistent with these previous findings. Although, if the Hopscotch is segregating as neutral variation in teosinte, the probability that it would have simultaneously drifted to high frequency in multiple populations is quite low; however, we find no evidence of selection or introgression at or near the *tb1* locus in our sample set. Interestingly, many of our populations with high frequency of the Hopscotch allele in the Jalisco region overlap with the Jalisco cluster in **(author?)** (Fukunaga).

many populations in the Jalisco cluster of Fukunaga seem to have high frequency of the hopscotch allele. Why? Only ancient introgression in Jalisco and not in the Balsas? If hopscotch came from standing variation, is this more consistent with a Jalisco domestication? I know, heresy!! LV, can you look at the parviglumis *populations with high hopscotch and see if they are all in the Jalisco cluster of Fukunaga? Might be a nice observation we could make here...could even go so far as saying the results from Matsuoka deserve another look. Matt, yes they are in the Jalisco cluster, I added in a sentence mentioning it, but am not sure if that's the direction you were wanting me to take with it or not*

The phenotypic effects of the Hopscotch insertion in domesticated maize have

been well documented (6; 45), and Weber et al. (51) have described its effects in partially inbred lines of teosinte (51). Our study is the first to explicitly examine the phenotypic effects of the Hopscotch insertion in individuals sampled from a natural population of teosinte. However, we found no significant effect of the Hopscotch on tillering index or tiller number in our phenotyping experiments, and the effect of the Hopscotch insertion in teosinte is discordant with that of maize. The lack of correlation between Hopscotch genotype and tillering index or tiller number is surprising given its effects in maize. *i took out the information about the TILS – seemed the consensus was more than we wanted to do for this pub. If qPCR is working for tb1 it would be something easy to check...* It is certainly possible that even though previous data demonstrate an effect of the Hopscotch on tillering in maize (45), that the effect of the Hopscotch in teosinte is more complicated and may be more difficult to observe. Moreover, *tb1* is a single gene in a complex pathway that affects branching and tillering traits, and perhaps in combination with alleles at other loci the phenotypic effects of the Hopscotch on tillering may not be consistent.

Variation at *tb1* has also been shown to contribute to phenotypes other than tillering (6), and a recent study by Studer and Doebley (44) examined the possibility of an allelic series at the *tb1* locus in teosinte. Studer and Doebley (44) introgressed 9 separate teosinte segments (one from *Zea diploperennis*, and four from both *Zea mays* ssp. *mexicana* and *Zea mays* ssp. *parviglumis*) spanning the *tb1* locus into an isogenic maize background and investigated their effects on previously associated phenotypes. They found that plants with teosinte chromosomal segments had greater tillering than their maize isogenic lines, and that different chromosomal segments of *tb1* confer different amounts of tillering, suggesting that there are multiple genetic factors in this region that affect tillering. However, in addition to elucidating variance in tillering among *tb1* teosinte segments, Studer and Doebley (44) found significant variance

among W22 control lines, suggesting that there are other genetic factors aside from alleles at the *tb1* locus that affect tillering in maize. Doebley and Stec (11, 12) first attempted to map QTL controlling many of the phenotypic differences between domesticated maize and teosinte, and demonstrated the existence of numerous QTL that contribute to the differences in branching architecture between the two. Many of these loci (*grassy tillers*, *gt1*; *tassel-replaces-upper-ears1*, *tru1*; *terminal ear1*, *ter1*) have been shown to interact with *tb1* (52; 32**?**), and both *tru1* and *ter1* have been shown to affect the same phenotypic traits as *tb1* (13). *tassel-replaces-upper-ears1* (*tru1*), for example, has been shown to act either epistatically or downstream of *tb1*, affecting both branching architecture (decreased apical dominance) and tassel phenotypes (shortened tassel and shank length and reduced tassel number) (32). It seems plausible that variation in some of these other loci could have affected tillering in our greenhouse population, and contributed to the lack of correlation we see between Hopscotch genotype and tillering.

In summary, our findings demonstrate that the Hopscotch allele is more widespread in populations of *parviglumis* and *mexicana* than previously thought. Analysis of linkage using SNPs from across chromosome 1 does not suggest that the Hopscotch allele is present in these populations due to recent introgression; however, it seems unlikely that the it would have drifted to high frequency in multiple populations, and there may be another explanation for the high frequency we observe in some of our populations. The Hopscotch does not appear to uniformly reduce tillering in teosinte as it does in maize, and the other loci involved in branching architecture of maize may play key roles in the regulation of tillering in teosinte. Furthermore, the high frequency with which we find the Hopscotch in a number of populations may suggest that the Hopscotch may play an additional ecological role in teosinte. In the future, additional experiments will be needed to examine expression levels of

tb1 and additional loci involved in branching architecture (e.g. *gt1*, *tru1*, and *ter1*) in conjuction with a more exhaustive phenotyping and genotyping assay.

seems like we should end with at least a hypothesis. as is we are saying we ruled out drift, introgression, and selection. what else is there? perhaps we should think about selection or introgression scenarios that could explain our data.

I'd lean toward saying that 1) hopscotch does not appear to uniformly reduce tillering in teosinte as it seems to in maize; 2) other loci are likely involved in regulating tillering; 3) given its high frequency in several teosinte populations, hopscotch may play additional ecological roles not obvious here based on what we chose to phenotype and, in the future, additional experiments are needed to look at 1) expression level differences at various loci shown to affect tillering in teosinte; and 2) to more completely phenotype teosinte plants with and without the hopscotch.

I went with Matt's suggestions. I am happy to include such scenarios, but everything I have thought of seems rather unlikely; i.e. only introgression or selection in Jalisco. Jeff I can mention that maybe parv is more of a weed there than other locations, but i thought that maize pollen spread rather substantial distances. If so wouldn't a number of our parv and mex sites would be within pollen dispersal distance?

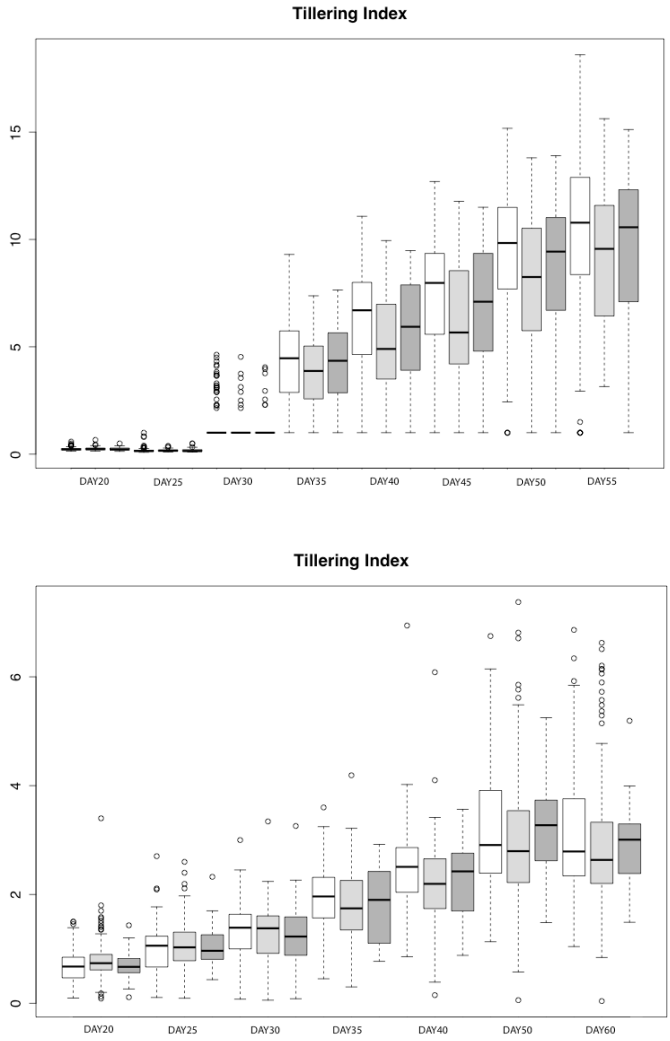


Figure 4: Box-plot showing tillering index in our greenhouse population from day 20-60. White indicates individuals homozygous for the Hopscotch, light grey represents heterozygotes, and dark grey represents homozygotes for the teosinte (No Hopscotch) allele. Within boxes, dark black lines represent the median, and the edges of the boxes are the first and third quartiles.

References

[1] Baack, E., Sapir, Y., Chapman, M., Burke, J., and Rieseberg, L. (2008). Selection on domestication traits and quantitative trait loci in crop-wild sunflower hybrids. *Mol Ecol*, 17(2):666–677.

[2] Barrett, S. (1983). Crop mimicry in weeds. *Econ Bot*, 37(3):255–282.

[3] Bradbury, P., Zhang, Z., Kroon, D., Casstevens, T., Ramdoss, Y., and Buckler, E. (2007). Tassel: software for association mapping of complex traits in diverse samples. *Bioinformatics*, 23(19):2633–2635.

[4] Briggs, W., McMullen, M., Gaut, B., and Doebley, J. (2007). Linkage mapping of domestication loci in a large maize-teosinte backcross resource. *Genetics*, 177(3):1915–1928.

[5] Chia, J., Song, C., Bradbury, P., Costich, D., de, Leon, N., Doebley, J., Elshire, R., Gaut, B., Geller, L., Glaubitz, J., Gore, M., Guill, K., Holland, J., Hufford, M., Lai, J., Li, M., Liu, X., Lu, Y., McCombie, R., Nelson, R., Poland, J., Prasanna, B., Pyhajarvi, T., Rong, T., Sekhon, R., Sun, Q., Tenaillon, M., Tian, F., Wang, J., Xu, X., Zhang, Z., Kaeppler, S., Ross-Ibarra, J., McMullen, M., Buckler, E., Zhang, G., Xu, Y., and Ware, D. (2012). Maize hapmap2 identifies extant variation from a genome in flux. *Nat Genet*, 44(7):803–U238.

[6] Clark, R., Wagler, T., Quijada, P., and Doebley, J. (2006). A distant upstream enhancer at the maize domestication gene tb1 has pleiotropic effects on plant and inflorescent architecture. *Nat Genet*, 38(5):594–597.

[7] Colosimo, P., Hosemann, K., Balabhadra, S., Villarreal, G., Dickson, M., Grimwood, J., Schmutz, J., Myers, R., Schluter, D., and Kingsley, D. (2005).

Widespread parallel evolution in sticklebacks by repeated fixation of ectodysplasin alleles. *Science*, 307(5717):1928–1933.

[8] Cook, J., McMullen, M., Holland, J., Tian, F., Bradbury, P., Ross-Ibarra, J., Buckler, E., and Flint-Garcia, S. (2012). Genetic architecture of maize kernel composition in the nested association mapping and inbred association panels. *Plant Physiol*, 158(2):824–834.

[9] Coop, G., Witonsky, D., Di, Rienzo, A., and Pritchard, J. (2010). Using environmental correlations to identify loci underlying local adaptation. *Genetics*, 185(4):1411–1423.

[10] Doebley, J. (2004). The genetics of maize evolution. *Annu Rev Genet*, 38:37–59.

[11] Doebley, J. and Stec, A. (1991). Genetic-analysis of the morphological differences between maize and teosinte. *Genetics*, 129(1):285–295.

[12] Doebley, J. and Stec, A. (1993). Inheritance of the morphological differences between maize and teosinte: Comparison of results for two f_2 populations. *Genetics*, 134(2):559–570.

[13] Doebley, J., Stec, A., and Gustus, C. (1995). *teosinte branched1* and the origin of maize: Evidence for epistasis and the evolution of dominance. *Genetics*, 141(1):333–346.

[14] Doebley, J., Stec, A., Wendel, J., and Edwards, M. (1990). Genetic and morphological analysis of a maize teosinte f2 population - implications for the origin of maize. *P Natl Acad Sci Usa*, 87(24):9888–9892.

[15] Doyle, J. and Doyle, J. (1990). A rapid total dna preparation procedure for small quantities of fresh tissue. *Phytochemical Bulletin*, 19:11–15.

[16] Ellstrand, N., Garner, L., Hedge, S., Guadagnuolo, R., and Blancas, L. (2007). Spontaneous hybridization between maize and teosinte. *Journal of Heredity*, 98(2):183–187.

[17] Ellstrand, N., Prentice, H., and Hancock, J. (1999). Gene flow and introgression from domesticated plants into their wild relatives. *Annu Rev Ecol Syst*, 30:539–563.

[18] Falush, D., Stephens, M., and Pritchard, J. (2003). Inference of population structure using multilocus genotype data: Linked loci and correlated allele frequencies. *Genetics*, 164(4):1567–1587.

[19] Feder, J., Berlocher, S., Roethele, J., Dambroski, H., Smith, J., Perry, W., Gavrilovic, V., Filchak, K., Rull, J., and Aluja, M. (2003). Allopatric genetic origins for sympatric host-plant shifts and race formation in rhagoletis. *P Natl Acad Sci Usa*, 100(18):10314–10319.

[20] Fox, J. and Weisberg, S. (2011). *An R Companion to Applied Regression*, volume Second Edition. Sage, Thousand Oaks, CA.

[Fukunaga] Fukunaga, K, N.-W. T. L. B. Z. Q. V. Y. F. M. B. K. L. L. D. J. Genetic diversity and population structure of teosinte. *Genetics*.

[22] Gallavotti, A., Zhao, Q., Kyozuka, J., Meeley, R., Ritter, M., Doebley, J., Pe, M., and Schmidt, R. (2004). The role of barren stalk1 in the architecture of maize. *Nature*, 432:630–635.

[23] Gerke, J., Edwards, J., KE, G., Ross-Ibarra, J., and McMullen, M. (2013). The genomic impacts of drift and selection for hybrid performance in maize. *arXiv*, 1307.7313.

[24] Goudet, J. (2005). Hierfstat, a package for r to compute and test hierarchical f-statistics. *Mol Ecol Notes*, 5(1):184–186.

[25] Hubner, S., Gunther, T., Flavell, A., Fridman, E., Graner, A., Korol, A., and Schmid, K. (2012). Islands and streams: clusters and gene flow in wild barley populations from the levant. *Mol Ecol*, 21(5):1115–1129.

[26] Hufford, M. (2010). Genetic and ecological approaches to guide conservation of teosinte (zea mays ssp. parviglumis), the wild progenitor of maize. *PhD Dissertation*, page 130pp.

[27] Hufford, M., Bilinski, P., Pyhjrvi, T., and Ross-Ibarra, J. (2012a). Teosinte as a model system for population and ecological genomics. *Trends in Genetics*, 12:606–615.

[28] Hufford, M., Lubinsky, P., Pyhjrvi, T., Devengenzo, M., Ellstrand, N., and Ross-Ibarra, J. (2013). The genomic signature of crop-wild introgression in maize. *PLoS Genetics*, 9(5):e1003477.

[29] Hufford, M., Xu, X., van, Heerwaarden, J., Pyhajarvi, T., Chia, J., Cartwright, R., Elshire, R., Glaubitz, J., Guill, K., Kaeppler, S., Lai, J., Morrell, P., Shannon, L., Song, C., Springer, N., Swanson-Wagner, R., Tiffin, P., Wang, J., Zhang, G., Doebley, J., McMullen, M., Ware, D., Buckler, E., Yang, S., and Ross-Ibarra, J. (2012b). Comparative population genomics of maize domestication and improvement. *Nat Genet*, 44(7):808–U118.

[30] Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P., and Drummond, A. (2012). Geneious basic: An integrated and

extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*, 28(12):1647–1649.

[31] Kitano, J., Bolnick, D., Beauchamp, D., Mazur, M., Mori, S., Nakano, T., and Peichel, C. (2008). Reverse evolution of armor plates in the threespine stickleback. *Curr Biol*, 18(10):769–774.

[32] Li, W. (2012). Tassels replace upper ears1 encodes a putative transcription factor that regulates maize shoot architecture by multiple pathways. *PhD Dissertation*, page 122.

[33] Lukens, L. and Doebley, J. (1999). Epistatic and environmental interactions for quantitative trait loci involved in maize evolution. *Genet Res*, 74(3):291–302.

[34] Maloof, M., Soliman, K., Jorgensen, R., and Allard, R. (1984). Ribosomal dna spacer length polymorphisms in barley - mendelian inheritance, chromosomal location, and population dynamics. *P Natl Acad Sci Usa*, 81:8014–8018.

[35] Matsuoka, Y., Vigouroux, Y., Goodman, M., Sanchez, G., Buckler, E., and Doebley, J. (2002). A single domestication for maize shown by multilocus microsatellite genotyping. *P Natl Acad Sci Usa*, 99(9):6080–6084.

[36] Nybom, H. (2004). Comparison of different nuclear dna markers for estimating intraspecific genetic diversity in plants. *Molecular Ecology*, 13:1143–1155.

[37] Olsen, K. and Gross, B. (2010). Genetic perspectives on crop domestication. *Trends in Plant Science*, 15(9):529–537.

[38] Piperno, D., Ranere, A., Holst, I., Iriarte, J., and Dickau, R. (2009). Starch grain and phytolith evidence for early ninth millennium bp maize from the central balsas river valley, mexico. *P Natl Acad Sci Usa*, 106(13):5019–5024.

[39] Plantinga, T., Alonso, S., Izagirre, N., Hervella, M., Fregel, R., van der Meer, J., Netea, M., and de la Rua, C. (2012). Low prevalence of lactase persistence in neolithic south-west europe. *Eur J Hum Genet*, 20(7):778–782.

[40] Pyhjrvi, T., Hufford, M., and Ross-Ibarra, J. (2013). Complex patterns of local adaptation in the wild relatives of maize. *Genome Biology and Evolution*, 5:1594–1609.

[41] Ross-Ibarra, J., Morrell, P., and Gaut, B. (2007). Plant domestication, a unique opportunity to identify the genetic basis of adaptation. *P Natl Acad Sci Usa*, 104(1):8641–8648.

[42] Ross-Ibarra, J., Tenaillon, M., and Gaut, B. (2009). Historical divergence and gene flow in the genus zea. *Genetics*, 181:1399–1413.

[43] Rozen, S. and Skaletsky, H. (2000). Primer3 on the www for general users and for biologist programmers. *Methods in Molecular Biology*, pages 365–386.

[44] Studer, A. and Doebley, J. (2012). Evidence for a natural allelic series at the maize domestication locus teosinte branched1. *Genetics*, 19(3):951–958.

[45] Studer, A., Zhao, Q., Ross-Ibarra, J., and Doebley, J. (2011). Identification of a functional transposon insertion in the maize domestication gene tb1. *Nat Genet*, 43(11):1160–U164.

[46] Thornton, K. (2003). libsequence: a c++ class library for evolutionary genetic analysis. *Bioinformatics*, 19(17):2325–2327.

[47] Thurber, C., Reagon, M., Gross, B., Olsen, K., Jia, Y., and Caicedo, A. (2010). Molecular evolution of shattering loci in us weedy rice. *Mol Ecol*, 19(16):3271–3284.

[48] Tishkoff, S., Reed, F., Ranciaro, A., Voight, B., Babbitt, C., Silverman, J., Powell, K., Mortensen, H., Hirbo, J., Osman, M., Ibrahim, M., Omar, S., Lema, G., Nyambo, T., Ghori, J., Bumpstead, S., Pritchard, J., Wray, G., and Deloukas, P. (2007). Convergent adaptation of human lactase persistence in africa and europe. *Nat Genet*, 39(1):31–40.

[49] van Heerwaarden, J., Doebley, J., Briggs, W., Glaubitz, J., Goodman, M., Gonzalez, J., and Ross-Ibarra, J. (2011). Genetic signals of origin, spread, and introgression in a large sample of maize landraces. *P Natl Acad Sci Usa*, 108(3):1088–1092.

[50] Wang, R., Stec, A., Hey, J., Lukens, L., and Doebley, J. (1999). The limits of selection during maize domestication. *Nature*, 398(6724):236–239.

[51] Weber, A., Clark, R., Vaughn, L., Sanchez-Gonzalez, J., Yu, J., Yandell, B., Bradbury, P., and Doebley, J. (2007). Major regulatory genes in maize contribute to standing variation in teosinte (zea mays ssp parviglumis). *Genetics*, 177(4):2349–2359.

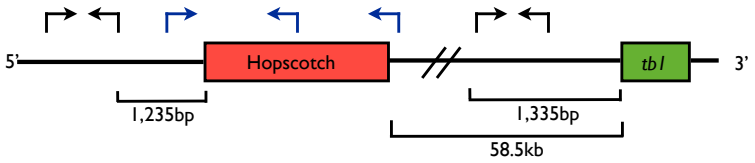
[52] Whipple, C., Kebrom, T., Weber, A., Yang, F., Hall, D., Meeley, R., Schmidt, R., Doebley, J., Brutnell, T., and Jackson, D. (2011). grassy tillers1 promotes apical dominance in maize and responds to shade signals in the grasses. *P Natl Acad Sci Usa*, 108(33):E506–E512.

[53] Wilkes, H. (1977). Hybridization of maize and teosinte, in mexico and guatemala and the improvement of maize. *Economic Botany*, 31:254–293.

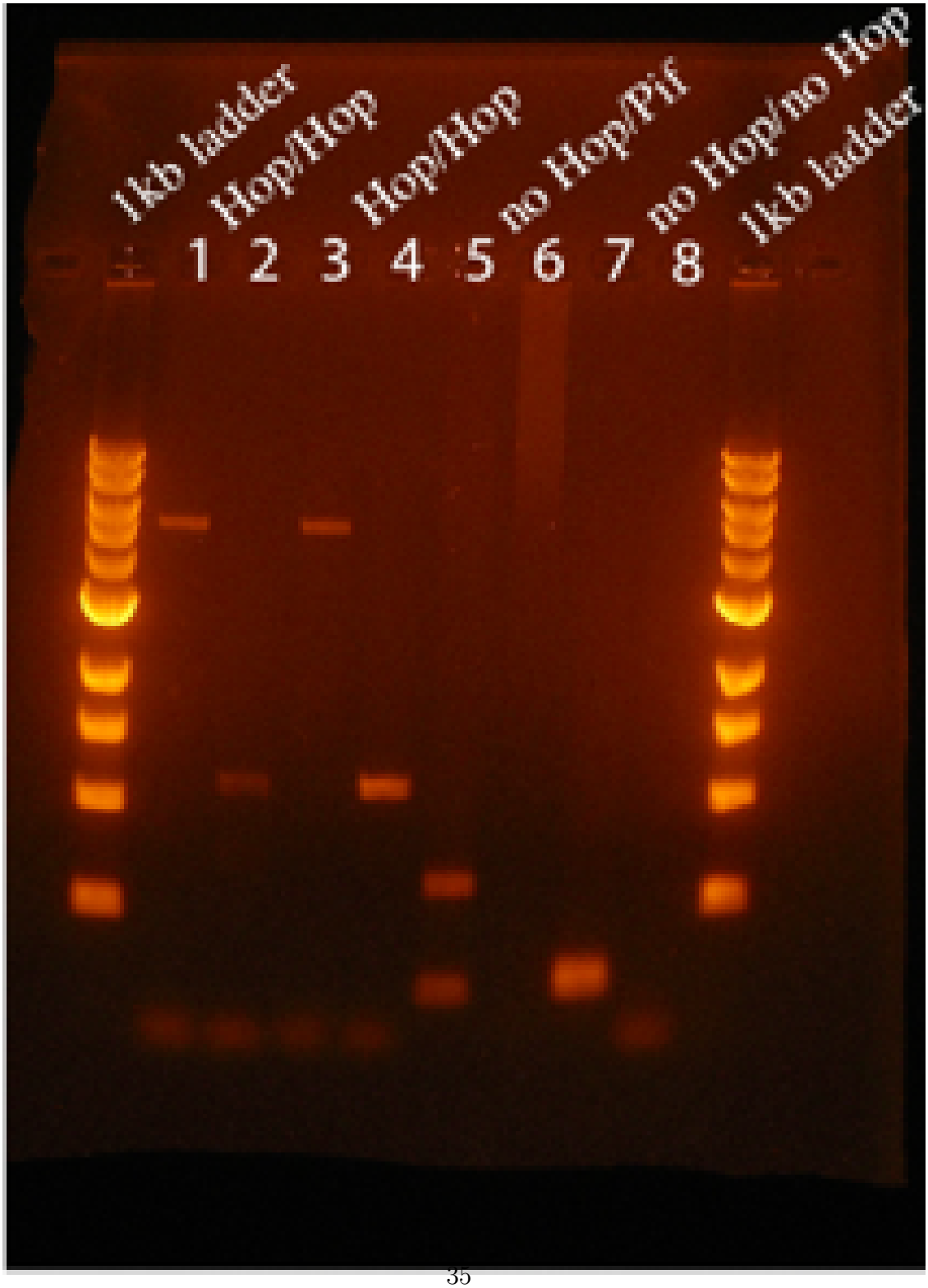
[54] Zhang, L., Zhu, Q., Wu, Z., Ross-Ibarra, J., Gaut, B., Ge, S., and Sang, T.

(2009). Selection on grain shattering genes and rates of rice domestication. *New Phytol*, 184(3):708–720.

Supplementary Materials



Sup.. 1: Representation of the upstream regulatory region of *tb1*, showing the *tb1* coding region (green) and the Hopscotch insertion (red). Arrows show the location of primer sets; in black, primers used for amplification and sequencing (Sequenced regions 1; within the 5? UTR, and sequenced region 2; 66,169 bp upstream from the tb1 ORF); in blue, primers used to genotype the Hopscotch insertion.



Sup.. 2: Agarose gel image of amplification products using our primer sets. Genotypes are indicated at the top of the gel.

Sup.. 3: Neighbor-Joining tree of the sequenced region in the 5? UTR (right; sequenced region 1) and the 66,169 bp upstream region (left; sequenced region 2) of *tb1*. Individuals with genotype data are colored: Homozygous for the teosinte (no Hopscotch) allele (red), homozygous for the maize (Hopscotch) allele (blue), heterozygotes (purple). TILs (teosinte inbred lines) are colored in green, with stars indicating the 3 TILs known to have the Hopscotch insertion. Black indicates indi-

Sup. 1: Accessions of *Zea mays* ssp. *mexicana* (RIMME) and *Zea mays* ssp. *parviglumis* (RIMPA) sampled. RIHY is a *Z. mays* ssp. *parviglumis* and *Zea mays* ssp. *mays* hybrid.

Accession	USDA Accession ID	Locality	Number alleles sampled	Hop
RIHY0009	N/A	N/A	2	
RIMME0006	566673	Durango, Mexico	2	
RIMME0007	566680	Guanajuato, Mexico	2	
RIMME0008	566681	Michoacan, Mexico	2	
RIMME0009	566682	Distrito Federal, Mexico	2	
RIMME0011	566685	Mexico, Mexico	2	
RIMME0014	714151	Breeders line; Puga: 11066	6	
RIMME0017	699874	Ayotlan, Mexico	8	
RIMME0021	N/A	El Porvenir, Mexico	69	
RIMME0026	N/A	Opopeo, Mexico	42	
RIMME0028	N/A	Puruandiro, Mexico	28	
RIMME0029	N/A	Ixtlan, Mexico	35	
RIMME0030	N/A	San Pedro, Mexico	27	
RIMME0031	N/A	Tenango del Aire, Mexico	25	
RIMME0032	N/A	Nabogame, Mexico	24	
RIMME0033	N/A	Puerta Encantada, Mexico	25	
RIMME0034	N/A	Santa Clara, Mexico	23	
RIMME0035	N/A	Xochimilco, Mexico	25	
RIMPA0001	87168	El Salado, Mexico	4	
RIMPA0003	87171	Mazatlan, Mexico	8	
RIMPA0017	87200	N/A	4	
RIMPA0019	87213	El Salado, Mexico	2	
RIMPA0029	87244	N/A	2	
RIMPA0031	87249	N/A	2	
RIMPA0035	87288	Jalisco, Mexico	4	
RIMPA0040	288185	Mexico, Mexico	4	
RIMPA0042	288187	Guerrero, Mexico	4	
RIMPA0043	288188	Guerrero, Mexico	4	

Sup. 2: *Zea mays* ssp. *mays* (RIMMA) sampled for genotyping

Accession	Number of alleles sampled	Hopscotch Frequency
RIMMA0066	2	1
RIMMA0075	2	1
RIMMA0077	2	1
RIMMA0079	2	1
RIMMA0081	2	1
RIMMA0084	2	1
RIMMA0086	2	1
RIMMA0088	2	1
RIMMA0089	2	1
RIMMA0090	2	1
RIMMA0092	4	1
RIMMA0094	4	1
RIMMA0097	2	1
RIMMA0099	2	1
RIMMA0100	2	1
RIMMA0101	2	1
RIMMA0104	2	1
RIMMA0108	2	1
RIMMA0111	6	1
RIMMA0115	2	1
RIMMA0117	2	1
RIMMA0130	2	1
RIMMA0133	2	1
RIMMA0134	2	1
RIMMA0135	2	1
RIMMA0142	2	0.5
RIMMA0143	4	1
RIMMA0146	4	1
RIMMA0149	2	1
RIMMA0152	2	1

Sup. 3: Variables and rotations used for the 6 principal components used for BayEnv calculations and their corresponding Bayes Factors. Modified from Ω ().

PC1		PC2		PC3		PC4		PC5	
Var	**Rot**	**Var**	**Rot**	**Var**	**Rot**	**Var**	**Rot**	**Var**	**Rot**
bio1	0.146	bio4	0.244	prec7	0.287	ts_clay	0.41	bio2	0.38
tmean11	0.146	bio3	0.241	prec8	0.276	v_mod	0.359	sq4	0.328
tmean12	0.145	bio7	0.241	prec11	0.262	ts_sand	0.329	ts_loam	0.289
bio11	0.145	prec6	0.237	bio13	0.247	bio15	0.272	ts_sand	0.266
tmax12	0.145	sq7	0.218	prec1	0.246	prec4	0.259	sq7	0.231
tmin5	0.145	prec9	0.217	bio16	0.242	x_mod	0.244	bio18	0.213
tmean1	0.145	sq3	0.207	prec12	0.24	prec3	0.226	bio13	0.207
tmean2	0.145	prec12	0.207	bio19	0.238	sq3	0.21	prec11	0.183
tmin4	0.145	bio12	0.204	bio12	0.231	prec5	0.21	bio7	0.17
tmax1	0.145	bio19	0.196	prec2	0.222	prec7	0.19	bio16	0.163
tmean4	0.145	prec2	0.188	bio18	0.221	sq4	0.186	bio4	0.157
tmin11	0.144	prec1	0.185	sq4	0.2	bio3	0.185	bio12	0.156
tmax11	0.144	prec10	0.184	prec9	0.18	bio18	0.178	bio3	0.155
tmin12	0.144	bio16	0.183	prec10	0.171	sq7	0.132	prec6	0.154
tmin2	0.144	prec8	0.17	prec5	0.161	bio14	0.116	x_mod	0.152
tmean5	0.144	prec5	0.165	prec4	0.154	bio13	0.099	prec9	0.144
tmean10	0.144	bio14	0.158	sq3	0.147	bio16	0.095	prec8	0.143
bio6	0.144	bio13	0.151	bio2	0.143	prec8	0.09	v_mod	0.142
tmax2	0.144	bio17	0.149	bio17	0.129	bio7	0.077	bio15	0.136
tmean3	0.144	prec3	0.144	ts_loam	0.127	bio4	0.075	prec7	0.112
tmin1	0.143	ts_clay	0.141	v_mod	0.123	bio2	0.074	prec4	0.108
tmin10	0.143	bio2	0.129	prec3	0.113	prec2	0.074	bio14	0.096
Altitude	0.143	prec7	0.108	x_mod	0.111	bio19	0.068	tmax7	0.093
bio9	0.143	tmax6	0.107	bio14	0.099	prec12	0.056	tmax8	0.092
tmin3	0.143	x_mod	0.106	bio4	0.07	ts_loam	0.053	prec1	0.091
bio10	0.142	bio15	0.098	tmax3	0.067	tmax12	0.047	prec2	0.086
tmax10	0.142	ts_loam	0.088	ts_clay	0.065	bio17	0.047	tmin11	0.086
tmax3	0.142	tmean6	0.085	bio15	0.056	bio9	0.043	prec5	0.082

Sup. 4: Diversity in the *tb1* region based on the maize SNP50 genotyping data

Population	**# seg sites**	θ_π **/ bp**	θ_W **/ bp**	**Tajima?s D**
Ejutla A	4	0.15217	0.11902	0.76191
Ejutla B	5	0.15258	0.14877	0.07412
La Mesa	3	0.12802	0.08926	1.09209
San Lorenzo	3	0.09098	0.08926	0.04845