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

Laura Vann[1], Thomas Konø[1], Tanja Pyhäjärvi[1], Matthew Hufford[1], and Jeffrey Ross-Ibarra[*1,2]

[1]Department of Plant Sciences, University of California Davis
[2]Center for Population Biology and Genome Center, University of California Davis

Please add addresses for all authors. All should have plant sciences, but Kono, Tanja, and Huff need their new addresses too.

ABSTRACT

The *teosinte branched1* (*tb1*) gene, a repressor of lateral organ growth, was identified as 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 *mexicana*. Analysis of linkage disequilibrium (LD) between the Hopscotch element and variation in surrounding regions does not support a hypothesis of recent introgression from maize into teosinte, yet we find no correlation between the frequency of the Hopscotch insertion and environmental variables. Additionally, we find no evidence that the Hopscotch plays

---

[*]rossibarra@ucdavis.edu

a similar role in branching architecture in teosinte as it does in domesticated maize.

1 INTRODUCTION

Domesticated crops and their wild progenitors provide an excellent system in which to study adaptation and genomic changes associated with human-mediated selection **?**. 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 **?**. 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 **??**. 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 **???**. 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 **??**, and genetic diversity and effective population size (Ne) of these taxa is estimated to be high **?** relative to other plants **?**.

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 **???**. 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 **?**. The effects of this insertion have been observed in tiller number in maize, but little is known about its role, if any, in teosinte **?**. 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 **?**. Furthermore, Studer and Doebley (2012) **?** 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. 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*.

2 METHODS

2.1 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 (Table S1, Table S2). DNA was extracted from leaf tissue using a modified CTAB approach **??**. We designed primers using PRIMER3 **?** implemented in Geneious **?** 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'-AACAGTATGATTTCATGGGACCG-3'; and HopIntR, 5'-CCTCCACCTCTCATGAGATCC-3' (Figure S1). 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 (Figure S2). 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.

2.2 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 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 **??**. We designed the following primers using PRIMER3 **?**: for the 5' UTR, 5' GGATAATGTGCACCAGGTGT 3' and 5' GCGTGC-TAGAGACACYTGTTGCT 3'; for the 50 kb upstream region, 5' TGTC-CTCGCCGCAACTC 3' and 5' TGTACGCCCGCCCCTCATCA 3' (Figure S1). 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 ex-

4

tended 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 ?. 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.

2.3   Genotyping Analysis

We examined discrepancies between observed and expected genotype frequencies by calculating Hardy-Weinberg Equilibrium (HWE), and to calculate differentiation between populations (FST) and subspecies (FCT) we used HierFstat ?. 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 ?. 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)?. We used genotyping and covariance data from Pyhjrvi et al., (2013) ? for BayEnv, with the Hopscotch insertion coded as an additional SNP (Table S3). Environmental data were obtained from www.worldclim.org, the Harmonized World Soil Database and www.harvestchoice.org, and summarized by principle component analy-

5

sis Pyhjrvi et al. (2013) **?**.

2.4 Sequence Analysis

For population genetic analyses of sequenced region 1 and sequenced region 2 we used the analysis package of Libsequence **?** to calculate pairwise FST 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, **?**).

In order to assess patterns of linkage disequilibrium (LD) around the Hopscotch element in the context of chromosomal patterns of LD we used Tassel **?** 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 **?**. 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 **?** 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 **?** and the same phased 55K SNP data from **?** that we used for LD analysis, combined with the corresponding SNP data from a diverse panel of 282 maize lines **?**. SNPs were anchored in a modified version of the IBM genetic map (**?**, `http://arxiv.org/abs/1307.7313`). We created haplotype blocks using a custom Perl script that grouped all SNPs within a 5 kb window into a haplotype. 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.

6

2.5 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) **?** 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 **?**. 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 **?**. 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 **?** 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.

3   Results

3.1   Genotyping

We genotyped 247 accessions of maize landraces, and found the Hopscotch element fixed in all but 8 of these accessions (inbred lines X,Y, Z, and landraces in Chihuahua, Tapalpa, Tenancingo, and Altotonga) (Table S1, Table S2). 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 Michoacn in central-western Mexico in both subspecies (Figure 1). 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 populations (FST) and subspecies (FCT) for populations in which we sampled 8 or more alleles. FCT is 0 within our dataset, and we found similar levels of FST between populations within each subspecies (0.22) and between all populations (0.23), to those reported in genome-wide estimates from previous studies **?** (Table 1). Although 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; Table S3).

3.2 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, Ejutla B, and La Mesa have comparable values for haplotype diversity,θ_W, θ_π, and Tajima's D, while San Lorenzo has lower values for haplotype diversity,θ_W, θ_π, and a more positive value for Tajima's D (Table 2). For region 2, haplotype diversity,θ_W, 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 San Lorenzo, indicating an excess of low frequency variants in this latter population (Table 2). Pairwise values of FST within population pairs Ejutla A/Ejutla B and San Lorenzo/La Mesa are 0 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, **?**) do not reveal any clear clustering pattern with respect to population or Hopscotch genotype (Figure S3). Furthermore, individuals within our sample that have the Hopscotch insertion do not group with the teosinte inbred lines or the lines of domesticated maize that have the Hopscotch insertion.

3.3 Evidence of introgression

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 (Figure 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 ($r^2 = 0.083$) (Table 3).

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 **??**. 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 *parviglumis* populations is lower (average 0.017) than the chromosome wide average (0.20). (Figure 3).

3.4 Phenotyping

3.4.1 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 (Figure 4). Additionally we find no significant correlation between tiller number and Hopscotch genotype, or culm diameter and

Hopscotch genotype in phenotyping 1.

3.5 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.

4 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 on lactase tolerance in humans **??**; variation at the Eda locus in three-spined stickleback **??**; and pupal diapause in the Apple Maggot fly **?**. 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, **?**; and grassy tillers1, **?** 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 **??**. Studer et al. (2011) genotyped 90 accessions of teosinte (inbred and non-inbred), and provided the first evidence that the Hopscotch insertion is segregating in teosinte **?**. Our findings suggest that the Hopscotch insertion is more widespread in populations of teosinte than previously thought and that the presence of the insertion is not due to recent introgression. However,

we find no effect of the Hopscotch insertion on branching architecture in teosinte.

Given that the Hopscotch insertion has been estimated to predate the domestication of maize, it is not surprising that it is still 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. (2011) in that we identified the Hopscotch allele segregating in teosinte, we find it at higher frequency than previously suggested **?**. Crop to wild introgression has previously been observed at domestication loci (**????????**), but our results are more consistent with Hufford et al. (2013) who found resistance to introgression from maize into teosinte **?**. Furthermore, Hufford et al. (2013) 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).

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. On the other hand, 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 (Figure 3, Table 4).

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 populations 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 **???**.

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 are consistent with patterns expected under neutrality. Though, 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. Similarly, 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.

The phenotypic effects of the Hopscotch insertion in domesticated maize have been well documented **??**, and Weber et al. (2007) have described its effects in partial inbred lines of teosinte **?**. Our study is the first to explicitly examine the phenotypic effects of the Hopscotch insertion in a 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, and preliminary data collected on a grow-out of the TILs (teosinte inbred lines).

It is certainly possible that even though data from the TILs suggest an effect of the Hopscotch on tillering in teosinte, that it in fact has no effect. Variation at *tb1* has also been shown to contribute to phenotypes other than tillering **?**, and perhaps we would have detected an effect of the Hopscotch on these additional phenotypes. A recent study by Studer and Doebley (2012) examined the possibility of an allelic series at the *tb1* locus in teosinte by introgressing 9 separate teosinte segments spanning the *tb1* locus into an isogenic maize background and investigating their effects on phenotype. However, they found evidence for an allelic series contributing to all *tb1* associated phenotypes but tillering. They did find variation in tillering both between introgressed lines and between isogenic maize lines, suggesting that there may be additional loci that contribute to tillering in both maize and teosinte **?**. 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.

Another explanation for the lack of correlation between Hopscotch genotype and tillering could be growing conditions in out greenhouse affected growth habit of our plants. The *tb1* gene is part of the shade avoidance

pathway in maize, and this involvement may have presented problems for our phenotyping design (**?**). Broadly, shade avoidance syndrome (SAS), refers to the morphological and physiological response of plants experiencing an increase in far-red (FR) light, which occurs during shading **?**. Under increased FR light conditions plants experience both morphological and physiological changes including reduction in branching, increased height, and increased ethylene and auxin production **??**. In maize, increase in shading has been shown to be correlated with changes in *tb1* expression and subsequent reduction in tillering, suggesting that *tb1* is involved in the shade avoidance response in maize **??**. If plants in our phenotyping experiments were grown too close together, such that they shaded each other, it could have reduced *tb1* expression and tillering, which would have confounded our results.

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 insertion does not appear to have any affect on tillering phenotypes in teosinte previously associated with *tb1* and it is possible that variation at other loci may play a larger role in branching architecture in teosinte than in domesticated maize.

BLAH (**?**) BLAH