Natural variation in teosinte at the domestication locus *teosinte branched1* (*tb1*)[1]

Laura Vann[2], Thomas Kono[2,3], Tanja Pyhäjärvi[2,4], Matthew B. Hufford[2,5,7], and Jeffrey Ross-Ibarra[2,6,7]

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

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

[4]Department of Biology, University of Oulu

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

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

[7]Author for Correspondence

1

Acknowledgements

1 Abstract

2 The *teosinte branched1* (*tb1*) gene, a repressor of lateral organ growth, is a major

3 QTL involved in branching differences between maize and its wild progenitor,

4 teosinte. Previous work has shown that the insertion of a transposable element

5 (*Hopscotch*) upstream of *tb1* enhances the gene's expression, causing much of the

6 reduction in tillering observed in domesticated maize. Observations of the maize

7 *tb1* allele in teosinte individuals, coupled with estimates of an age of insertion of

8 the *Hopscotch* element that predates domestication, led us to investigate its

9 prevalence and potential role in teosinte. Results from genotyping across many

10 natural populations suggest that the *Hopscotch* element is segregating at a higher

11 than expected frequency in a number of populations of two subspecies of teosinte,

12 *Zea mays* ssp. *parviglumis* and *Zea mays* ssp. *mexicana*. Analysis of linkage

13 disequilibrium between the *Hopscotch* element and variation in surrounding regions

14 does not support a hypothesis of recent introgression from maize into teosinte.

15 Population genetic signatures, however, are consistent with selection on this locus

16 and suggest the *Hopscotch* insertion at *tb1* may play an ecological role in teosinte.

17 Finally, two greenhouse experiments with *Zea mays* ssp. *parviglumis* do not suggest

18 *tb1* controls tillering in natural populations of this subspecies. Our findings suggest

19 that the role of the *Hopscotch* in tillering in teosinte differs from domesticated

20 maize, and that other loci may play a role in observed variation in this trait.

INTRODUCTION

2 Domesticated crops and their wild progenitors provide an excellent system in which

3 to study adaptation and genomic changes associated with human-mediated

4 selection (Ross-Ibarra et al., 2007). Perhaps the central focus of the study of

5 domestication has been the identification of genetic variation underlying

6 agronomically important traits such as fruit size and plant architecture (Olsen and

7 Gross, 2010). Additionally, many domesticates show reduced genetic diversity when

8 compared to their wild progenitors, and an understanding of the distribution of

9 diversity in the wild and its phenotypic effects has become increasingly useful to

10 crop improvement (Kovach and McCouch, 2008). But while some effort has been

11 invested into understanding how wild alleles behave in their domesticated relatives

12 (e.g. Bai and Lindhout, 2007), very little is known about the role that alleles found

13 most commonly in domesticates play in natural populations of their wild

14 progenitors (Whitton J, 1997).

15 Maize (*Zea mays* ssp. *mays*) was domesticated from the teosinte *Zea mays* ssp.

16 *parviglumis* (hereafter, *parviglumis*) roughly 9,000 B.P. in southwest Mexico

17 (Piperno et al., 2009; Matsuoka et al., 2002). Domesticated maize and the teosintes

18 are an attractive system in which to study domestication due to the abundance of

19 genetic tools developed for maize and well-characterized domestication loci (Hufford

20 et al., 2012a; Doebley, 2004; Hufford et al., 2012b). Additionally, large naturally

21 occurring populations of both *Zea mays* ssp. *parviglumis* (the wild progenitor of

22 maize) and *Zea mays* ssp. *mexicana* (highland teosinte; hereafter *mexicana*) can be

23 found throughout Mexico (Wilkes, 1977; Hufford et al., 2013), and genetic diversity

24 of these taxa is estimated to be high (Ross-Ibarra et al., 2009).

25 Many morphological changes are associated with maize domestication, and

26 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 ancient 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 (Studer and Doebley, 2012), but little is known about the role of *tb1* or the *Hopscotch* insertion in natural populations of teosinte.

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* and other grasses and may therefore be crucial to the ecology of teosinte (Kebrom and Brutnell, 2007; Lukens and Doebley, 1999). In this study

1   we aim to characterize the distribution of the *Hopscotch* insertion in *parviglumis*,

2   *mexicana*, and landrace maize, and to examine the phenotypic effects of the

3   insertion in *parviglumis*. We use a combination of PCR genotyping for the

4   *Hopscotch* element in our full panel and sequencing of two small regions upstream

5   of *tb1* in a subset of teosinte populations to explore patterns of genetic variation at

6   this locus. Finally, we test for an association between the *Hopscotch* element and

7   tillering phenotypes in a population of *parviglumis*.

8   <h1 style="text-align:center">MATERIALS AND METHODS</h1>

9   **Sampling and genotyping**

10   We sampled 1,110 individuals from 350 accessions (247 maize landraces, 17

11   *mexicana* populations, and 86 *parviglumis* populations) and assessed the presence

12   or absence of the *Hopscotch* insertion (Table S1 and Table S2). DNA was extracted

13   from leaf tissue using a modified CTAB approach (Doyle and Doyle, 1990; Maloof

14   et al., 1984). We designed primers using PRIMER3 (Rozen and Skaletsky, 2000)

15   implemented in Geneious (Kearse et al., 2012) to amplify the entire *Hopscotch*

16   element, as well as an internal primer allowing us to simultaneously check for

17   possible PCR bias between presence and absence of the *Hopscotch* insertion. Two

18   PCRs were performed for each individual, one with primers flanking the *Hopscotch*

19   (HopF/HopR) and one with a flanking primer and an internal primer

20   (HopF/HopIntR). Primer sequences are HopF,

21   5'-TCGTTGATGCTTTGATGGATGG-3'; HopR, 5'-AACAGTATGATTTCATGGGACCG-3';

22   and HopIntR, 5'-CCTCCACCTCTCATGAGATCC-3' (Fig. S1, Fig. S2) *Primers in Fig. S1*

23   *should be labeled* . Homozygotes show a single band for absence of the element

24   (∼300bp) and two bands for presence of the element (∼5kb and XX *LV, please add the*

1 *size of the second band*), whereas heterozygotes are three-banded (Fig. S2). When only

2 one PCR resolved well, we scored one allele for the individual. We used Phusion

3 High Fidelity Enzyme (Thermo Fisher Scientific Inc., Waltham, Massachusetts,

4 USA) and the following conditions for amplifications: 98°C for 3 min, 30 cycles of

5 98°C for 15 s, 65°C for 30 s, and 72°C for 3 min 30 s, with a final extension of 72°C

6 for 10 min. PCR products were visualized on a 1% agarose gel and scored for

7 presence/absence of the *Hopscotch* based on band size.

8 Sequencing

9 In addition to genotyping, we chose a subset of *parviglumis* individuals for

10 sequencing. We chose twelve individuals from each of four populations from Jalisco

11 state, Mexico (San Lorenzo, La Mesa, Ejutla A, and Ejutla B). For amplification

12 and sequencing, we selected two regions approximately 600bp in size from within

13 the 5' UTR of *tb1* (Region 1) and from 1,235bp upstream of the start of the

14 *Hopscotch* (66,169bp upstream from the start of the *tb1* ORF; Region 2). We

15 designed the following primers using PRIMER3 (Rozen and Skaletsky, 2000): for

16 the 5' UTR, 5'-GGATAATGTGCACCAGGTGT-3' and

17 5'-GCGTGCTAGAGACACYTGTTGCT-3'; for the 66kb upstream region,

18 5'-TGTCCTCGCCGCAACTC-3' and 5'-TGTACGCCCGCCCCTCATCA-3' (Fig. S1). We

19 used Taq polymerase (New England Biolabs Inc., Ipswich, Massachusetts, USA)

20 and the following thermal cycler conditions to amplify fragments: 94°C for 3 min,

21 30 cycles of 92°C for 40 s, annealing for 1 min, 72°C for 40 s, and a final 10 min

22 extension at 72°C. Annealing temperatures for Region 1 and Region 2 were 59.7°C

23 and 58.8°C, respectively. To clean excess primer and dNTPs we added two units of

24 Exonuclease1 and 2.5 units of Antarctic Phosphatase to 8.0 μL of amplification

25 product. This mix was placed on a thermal cycler with the following program:

1   37°C for 30 min, 80°C for 15 min, and a final cool-down step to 4°C.

2   We cloned cleaned fragments into a TOPO-TA vector (Life Technologies,

3   Grand Island, New York, USA) using OneShot TOP10 chemically competent *E.*

4   *coli* cells, with an extended ligation time of 30 min for a complex target fragment.

5   We plated cells on LB agar plates containing kanamycin, and screened colonies

6   using vector primers M13 Forward and M13 Reverse under the following conditions:

7   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

8   a final extension at 72°C for 4 min. We visualized amplification products for

9   incorporation of our insert on a 1% agarose TAE gel.

10   Amplification products with successful incorporation of our insert were cleaned

11   using Exonuclease 1 and Antarctic Phosphatase following the procedures detailed

12   above, and sequenced with vector primers M13 Forward and M13 Reverse using

13   Sanger sequencing at the College of Agriculture and Environmental Sciences

14   (CAES) sequencing center at UC Davis. We aligned and trimmed primer sequences

15   from resulting sequences using the software Geneious (Kearse et al., 2012).

16   Following alignment, we verified singleton SNPs by sequencing an additional one to

17   four colonies from each clone. If the singleton was not present in these additional

18   sequences it was considered an amplification or cloning error, and we replaced the

19   base with the base of the additional sequences. If the singleton appeared in at least

20   one of the additional sequences we considered it a real variant and kept it for

21   further analyses.

Genotyping analysis

23   We examined discrepancies between observed and expected genotype frequencies by

24   calculating Hardy-Weinberg Equilibrium (HWE). To calculate differentiation

25   between populations (F_{ST}) and subspecies (F_{CT}) we used HierFstat (Goudet,

1 2005). These analyses only included populations in which 8 or more individuals

2 were sampled. To test the hypothesis that the *Hopscotch* insertion may be adaptive

3 under certain environmental conditions, we looked for significant associations

4 between the *Hopscotch* frequency and environmental variables using BayEnv (Coop

5 et al., 2010). BayEnv creates a covariance matrix of relatedness between

6 populations and then tests a null model that allele frequencies in populations are

7 determined by the covariance matrix of relatedness alone against the alternative

8 model that allele frequencies are determined by a combination of the covariance

9 matrix and an environmental variable, producing a posterior probability (*i.e.,*

10 Bayes Factor; Coop et al. 2010). We used genotyping and covariance data from

11 Pyhäjärvi et al. (2013) for BayEnv, with the *Hopscotch* insertion coded as an

12 additional SNP (Table S3). Environmental data were obtained from

13 `www.worldclim.org`, the Harmonized World Soil Database

14 (FAO/IIASA/ISRIC/ISSCAS/JRC, 2012) and `www.harvestchoice.org` and

15 summarized by principle component analysis following Pyhäjärvi et al. (2013).

16 **Sequence analysis**

17 For population genetic analyses of sequenced Region 1 and sequenced Region 2 we

18 used the Libsequence package (Thornton, 2003) to calculate pairwise F_{ST} between

19 populations and to calculate standard diversity statistics (number of haplotypes,

20 haplotype diversity, Watterson's estimator $\hat{\theta}_W$, pairwise nucleotide diversity $\hat{\theta}_\pi$, and

21 Tajima's D). To produce a visual representation of differentiation between

22 sequences and examine patterns in sequence clustering by *Hopscotch* genotype we

23 used Phylip (`http://evolution.genetics.washington.edu/phylip.html`),

24 creating neighbor-joining trees with bootstrap-supported nodes (100 repetitions).

25 For creation of trees we also included homologous sequence data from Maize

9

1 HapMapV2 (Chia et al., 2012) for teosinte inbred lines (TILs), some of which are

2 known to be homozygous for the *Hopscotch* insertion (TIL03, TIL17, TIL09), as

3 well as 59 lines of domesticated maize.

Introgression analysis

5 In order to assess patterns of linkage disequilibrium (LD) around the *Hopscotch*

6 element in the context of chromosomal patterns of LD we used Tassel (Bradbury

7 et al., 2007) and calculated LD between SNPs across chromosome 1 using

8 previously published data from twelve plants each of the Ejutla A (EjuA), Ejutla B

9 (EjuB), San Lorenzo (SLO), and La Mesa (MSA) populations (Pyhäjärvi et al.,

10 2013). We chose these populations because we had both genotyping data for the

11 *Hopscotch* as well as chromosome-wide SNP data for chromosome 1. For each

12 population we filtered the initial set of 5,897 SNPs on chromosome 1 to accept only

13 SNPs with a minor allele frequency of at least 0.1, resulting in 1,671, 3,023, 3,122,

14 and 2,167 SNPs for SLO, EjuB, EjuA, and MSA, respectively. We then used Tassel

15 (Bradbury et al., 2007) to calculate linkage disequilibrium (r^2) across chromosome

16 1 for each population.

17 We examined evidence of introgression on chromosome 1 in these same four

18 populations (EjuA, EjuB, MSA, SLO) using STRUCTURE (Falush et al., 2003)

19 and phased data from Pyhäjärvi et al. (2013), combined with the corresponding

20 SNP data from a diverse panel of 282 maize lines (Cook et al., 2012). SNPs were

21 anchored in a modified version of the IBM genetic map (Gerke et al., 2013). We

22 created haplotype blocks using a custom Perl script that grouped SNPs separated

23 by less than 5kb into haplotypes. We ran STRUCTURE at K=2 under the linkage

24 model, performing 3 replicates with an MCMC burn-in of 10,000 steps and 50,000

25 steps post burn-in.

1 Phenotyping of *parviglumis*

2 To investigate the phenotypic effects of the *Hopscotch* insertion in teosinte, we

3 conducted an initial phenotyping trial (Phenotyping 1). We germinated 250 seeds

4 of *parviglumis* collected in Jalisco state, Mexico (population San Lorenzo) (Hufford,

5 2010) where the *Hopscotch* is segregating at highest frequency (0.44) in our initial

6 genotyping sample set. In order to maximize the likelihood of finding the *Hopscotch*

7 in our association population we selected seeds from sites where genotyped

8 individuals were homozygous or heterozygous for the insertion. We chose between

9 10-13 seeds from each of 23 sampling sites. We treated seeds with fungicide and

10 germinated them in petri dishes with filter paper. Following germination, 206

11 successful germinations were planted into one-gallon pots with potting soil and

12 randomly spaced one foot apart on greenhouse benches. Plants were watered three

13 times a day by hand and with an automatic drip containing 10-20-10 fertilizer.

14   Starting on day 15, we measured tillering index as the ratio of the sum of tiller

15 lengths to the height of the plant (Briggs et al., 2007). Following initial

16 measurements, we phenotyped plants for tillering index every 5 days through day

17 40, and then on day 50 and day 60. On day 65 we measured culm diameter

18 between the third and fourth nodes of each plant. Culm diameter is not believed to

19 be correlated with tillering index or variation at *tb1*. Following phenotyping we

20 extracted DNA from all plants using a modified SDS extraction protocol

21 (`http://www.ars.usda.gov`). *what is this url?* We genotyped individuals for the

22 *Hopscotch* insertion following the protocols listed above. Based on these initial

23 data, we conducted a *post hoc* power analysis using data from day 40 of

24 Phenotyping 1, indicating that a minimum of 71 individuals in each genotypic class

25 would be needed to detect the observed effect of the *Hopscotch* on tillering index.

26   We performed a second phenotyping experiment (Phenotyping 2) in which we

1 germinated 372 seeds of *parviglumis*, choosing equally between sites previously

2 determined to have or not have the *Hopscotch* insertion. Seeds were germinated

3 and planted on day 7 post fruit-case removal into two gallon pots. Plants were

4 watered twice daily, alternating between fertilized and non-fertilized water. We

5 began phenotyping successful germinations (302 plants) for tillering index on day

6 15 post fruit-case removal, and phenotyped every five days until day 50. At day 50

7 we measured culm diameter between the third and fourth nodes. We extracted

8 DNA and genotyped plants following the same guidelines as in Phenotyping 1.

9   Tillering index data for each genotypic class did not meet the criteria for a

10 repeated measures ANOVA, so we transformed the data using a Box-Cox

11 transformation ($\alpha = 0$ *what is the alpha value here?* ; Car Package for R, Fox and Weisberg

12 2011) to improve the normality and homogeneity of variance among genotype

13 classes. We analyzed relationships between genotype and tillering index and tiller

14 number using a repeated measures ANOVA through a general linear model function

15 implemented in SAS v.9.3 (SAS Institute Inc., Cary, NC, USA). Additionally, in

16 order to compare any association between *Hopscotch* genotype and tillering and

17 associations at other presumably unrelated traits, we performed an ANOVA

18 between culm diameter and genotype using the same general linear model in SAS.

RESULTS

Genotyping

21 Genotype of the *Hopscotch* insertion was confirmed with two PCRs for 837

22 individuals. Among the 247 maize landrace accessions genotyped, all but eight were

23 homozygous for the presence of the insertion (Table S1 and Table S2). Within our

24 *parviglumis* and *mexicana* samples we found the *Hopscotch* insertion segregating in

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 alleles sampled.
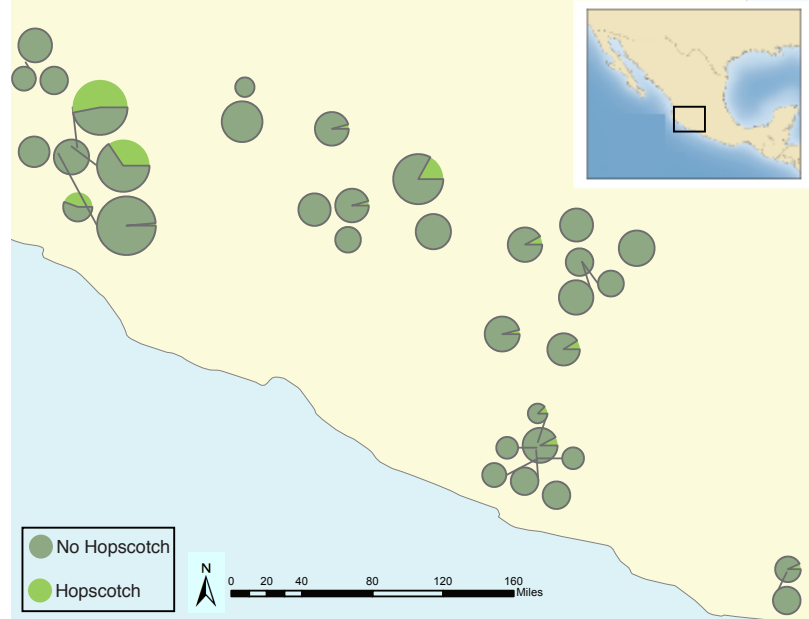
1 37 and 4 populations, respectively, and at highest frequency in the states of Jalisco,

2 Colima, and Michoacán in central-western Mexico (Fig. 1).

3 Using our *Hopscotch* genotyping, we calculated differentiation between

4 populations (F_{ST}) and subspecies (F_{CT}) for populations in which we sampled 8 or

5 more alleles. We found that $F_{CT} = 0$, and levels of F_{ST} among populations within

6 each subspecies (0.22) and among all populations (0.23) *are these an average of pairwise or is*

7 *this calculated among all pops?* are similar to those reported genome-wide in previous

8 studies (Pyhäjärvi et al. 2013; Table 1). Although we found large variation in

9 *Hopscotch* allele frequency among our populations, BayEnv analysis did not

10 indicate a correlation between the *Hopscotch* insertion and environmental variables

11 (all Bayes Factors < 1; Table S3).

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

Comparison	Region 1	Region 2	*Hopscotch*
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

1 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 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 less negative in San Lorenzo (Table 2). *need to reference* Hopscotch *frequencies in supplemental table somewhere* 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 San Lorenzo, *is the table wrong? MSA is the only negative value in the table* 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 0 for both sequenced

1 regions as well as for the *Hopscotch* table 1 shows 0.016 for hopscotch, not 0. which is right? ,

2 while they are high for other population pairs (Table 1). Neighbor joining trees of

3 our sequence data and data from the teosinte inbred lines (TILs; data from Maize

4 HapMapV2, Chia et al. 2012) do not reveal any clear clustering pattern with

5 respect to population or *Hopscotch* genotype (Figure S3); individuals within our

6 sample that have the *Hopscotch* insertion do not group with the teosinte inbred

7 lines or domesticated maize that have the *Hopscotch* insertion.

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

Population	# Haplotypes	Hap. Diversity	$\hat{\theta}_\pi$	Tajima's D
Region 1(5' UTR)				
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
Region 2 (66kb upstream)				
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

8 Evidence of introgression

9 The highest frequency of the *Hopscotch* insertion in teosinte was found in

10 *parviglumis* sympatric with cultivated maize. Our initial hypothesis was that the

11 high frequency of the *Hopscotch* element in these populations could be attributed

1 to introgression from maize into teosinte. To investigate this possibility we

2 examined overall patterns of linkage disequilibrium across chromosome one and

3 specifically in the *tb1* region. If the *Hopscotch* is found in these populations due to

4 recent introgression we would expect to find large blocks of linked markers near this

5 element. We find no evidence of elevated linkage disequilibrium between the

6 *Hopscotch* and SNPs surrounding the *tb1* region in our resequenced populations

7 (Fig. 2), and r^2 in the *tb1* region does not differ significantly between populations

8 with (average r^2 of 0.085) and without (average $r^2 = 0.082$) the *Hopscotch*

9 genotype. In fact, average r^2 is lower in the *tb1* region ($r^2 = 0.056$) than across the

10 rest of chromosome 1 ($r^2 = 0.083$) (3). *LV, please go through and make sure the data entered into*

11 *all the tables is correct. In Table3, both sequenced regions were labeled as "Region 1". I changed the second to*

12 *Region 2 but don't know if the data in this column are really from Region 2*

Table 3: 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	*tb1* 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

13   The lack of clustering of *Hopscotch* genotypes in our NJ tree as well as the lack

14 of LD around *tb1* does not support the hypothesis that the *Hopscotch* insertion in

15 these populations of *parviglumis* is the result of recent introgression. However, to

16 further explore this hypothesis we performed a STRUCTURE analysis using

17 Illumina MaizeSNP50 data from four of our *parviglumis* populations (EjuA, EjuB,
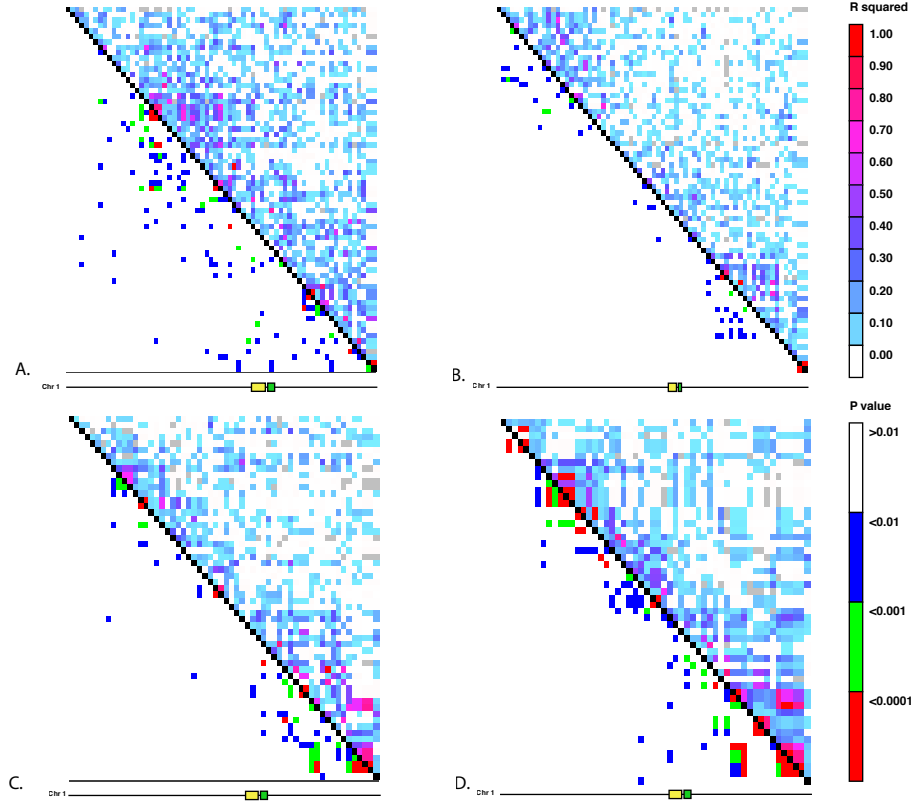
16

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 represents the *tb1* ORF. A) Ejutla A; B) Ejutla B; C) La Mesa; D). San Lorenzo

this needs description of what upper and lower triangle are

1 MSA, and SLO) and the maize 282 diversity panel (Cook et al., 2012; Pyhäjärvi

2 et al., 2013). The linkage model implemented in STRUCTURE can be used to

3 identify ancestry of blocks of linked variants, which would arise as a result of recent

4 admixture between populations. If the *Hopscotch* insertion is present in

5 populations of *parviglumis* as a result of recent admixture with domesticated maize,

6 we would expect the insertion and linked variants in surrounding sites to be

1 assigned to the "maize" cluster in our STRUCTURE runs, not the "teosinte"

2 cluster. In all runs, assignment to maize in the *tb1* region across all four

3 *parviglumis* populations is low (average 0.017) and much below the

4 chromosome-wide average (0.20; Table 4; Fig. 3).

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

	tb1 region		Chr 1	
Population	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

5 Phenotyping

6 To assess the contribution of *tb1* to phenotypic variation in tillering in a natural

7 population, we grew plants from seed sampled from the San Lorenzo population of

8 *parviglumis*, which had a high mean frequency (0.44) of the *Hopscotch* insertion

9 based on our initial genotyping. We measured tillering index (TI), the ratio of the

10 sum of tiller lengths to plant height, for 216 plants (Phenotyping 1) from within

11 the San Lorenzo population, and genotyped plants for the *Hopscotch* insertion. We

12 found the *Hopscotch* segregating at a frequency of 0.65 with no significant

13 deviations from expected frequencies under Hardy-Weinberg equilibrium. After

14 performing a repeated measures ANOVA between our transformed tillering index

15 data and *Hopscotch* genotype we find a weak positive correlation between presence

16 of the *Hopscotch* and tillering index on day 40 (p=0.0848), a result indicating the
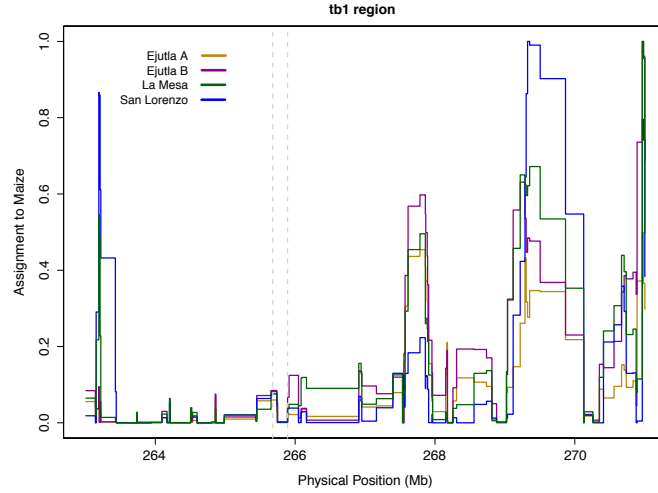
Figure 3: STRUCTURE assignment to maize across a section of chromosome 1. The dotted lines mark the beginning of the sequenced region 50kb upstream (Sequenced region 2) and the end of the *tb1* ORF.

1 *Hopscotch* may actually increase tillering in *parviglumis* in contrast to its

2 phenotypic effect in maize. We find no correlation between tillering index and

3 genotype on any other day (4). Additionally we find no significant correlation

4 between tiller number and *Hopscotch* genotype, or culm diameter and *Hopscotch*

5 genotype in Phenotyping 1.

6   We performed a second grow-out of *parviglumis* from San Lorenzo

7 (Phenotyping 2) to assess whether lighting conditions or sample size may have

8 affected our ability to detect an effect of *tb1*. For the second grow-out we measured

9 tillering index every five days through day 50 for 302 plants. We found the

10 *Hopscotch* allele segregating at a frequency of 0.69, *is it in HWE in this pop?*   with a 0.6

11 frequency of *Hopscotch* homozygotes, and a 0.2 frequency of both heterozygotes

12 and homozygotes for the teosinte allele. We found similar patterns, with a weak

19
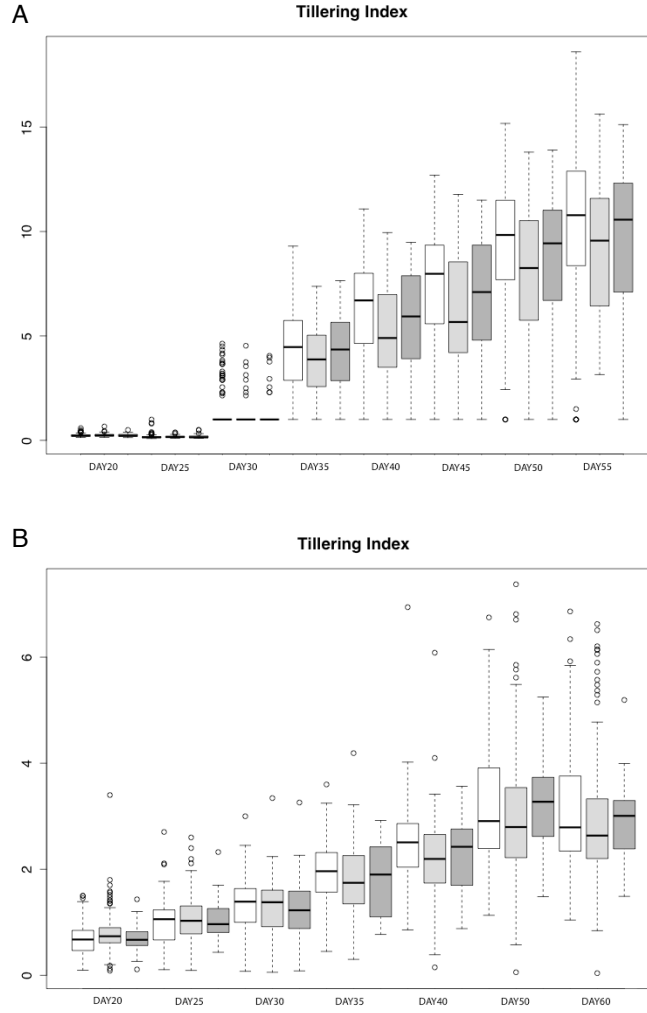
Figure 4: Box-plots showing tillering index in greenhouse grow-outs for Phenotyping 1 (A) and Phenotyping 2 (B). White indicates individuals homozygous for the *Hopscotch*, light grey represents heterozygotes, and dark grey represents homozygotes for the teosinte (No *Hopscotch*) allele. Within boxes, dark black lines represent the median, and the edges of the boxes are the first and third quartiles. *please explain whiskers and dots on figure too.*

20

1 positive correlation between tillering index and *Hopscotch* genotype at day 40

2 (p<0.0611), with no significant correlation on any day. Similarly, relationships

3 between *Hopscotch* genotype and tiller number and *Hopscotch* genotype and culm

4 diameter were not significant.

DISCUSSION

6   Adaptation occurs due to selection on standing variation or *de novo* mutations.

7 Adaptation from standing variation has been well-described in a number of systems;

8 for example, selection for lactose tolerance in humans (Plantinga et al., 2012;

9 Tishkoff et al., 2007), variation at the *Eda* locus in three-spined stickleback (Kitano

10 et al., 2008; Colosimo et al., 2005), and pupal diapause in the Apple Maggot fly

11 (Feder et al., 2003). Although the adaptive role of standing variation has been

12 described in many systems, its importance in domestication is not as well studied.

13   In maize, alleles at domestication loci (*RAMOSA1*, Sigmon and Vollbrecht

14 2010; *barren stalk1*, Gallavotti et al. 2004; and *grassy tillers1*, Whipple et al. 2011)

15 are thought to have been selected from standing variation, suggesting that diversity

16 already present in teosinte may have played an important role in maize

17 domestication. The *teosinte branched1* gene is one of the best characterized

18 domestication loci, and, while previous studies have suggested that differences in

19 plant architecture between maize and teosinte are a result of selection on standing

20 variation at this locus, little is known about natural variation at this locus and its

21 ecological role in teosinte (Clark et al., 2006; Studer et al., 2011). Studer et al.

22 (2011) genotyped 90 accessions of teosinte (inbred and outbred), providing the first

23 evidence that the *Hopscotch* insertion is segregating in teosinte (Studer et al., 2011).

24   Given that the *Hopscotch* insertion has been estimated to predate the

25 domestication of maize, it is not surprising that it can be found segregating in

21

1 populations of teosinte. However, by widely sampling across teosinte populations

2 our study provides greater insight into the distribution and prevalence of the

3 *Hopscotch* in teosinte. While our findings are consistent with Studer et al. (2011) in

4 that we identify the *Hopscotch* allele segregating in teosinte, we find it at higher

5 frequency than previously suggested (Studer et al., 2011). Many of our populations

6 with high frequency of the *Hopscotch* allele fall in the Jalisco cluster identified by

7 Fukunaga (2005), suggesting a different history of the *tb1* locus in this region than

8 in the Balsas River Basin where maize was domesticated (Matsuoka et al., 2002).

9 Potential explanations for the high frequency of the *Hopscotch* element in

10 *parviglumis* from the Jalisco cluster include gene flow from maize, genetic drift, and

11 natural selection.

12 While gene flow from crops into their wild relatives is well-known, (Ellstrand

13 et al., 1999; Zhang et al., 2009; Thurber et al., 2010; Baack et al., 2008; Hubner

14 et al., 2012; Wilkes, 1977; van Heerwaarden et al., 2011; Barrett, 1983), our results

15 are more consistent with Hufford et al. (2013) who found resistance to introgression

16 from maize into teosinte. Furthermore, Hufford et al. (2013) showed that

17 domestication loci, such as *tb1*, are particularly resistant to introgression in both

18 directions of gene flow (i.e., maize to teosinte and teosinte to maize). We find no

19 evidence of recent introgression in our analyses. Clustering patterns in our NJ trees

20 do not reflect a pattern expected if maize alleles at the *tb1* locus had introgressed

21 into populations of teosinte. Moreover, there is no signature of elevated LD in the

22 *tb1* region relative to the rest of chromosome 1, and Bayesian assignment to a maize

23 cluster in this region is lower than the chromosome-wide average (Fig. 3, Table 4).

24 Together, these data point to an explanation other than recent introgression for the

25 high observed frequency of *Hopscotch* in a subset of our *parviglumis* populations.

26 Although recent introgression seems unlikely, we cannot rule out ancient

22

introgression as an explanation for the presence of the *Hopscotch* in these populations. If the *Hopscotch* allele was introgressed in the distant past, recombination may have broken up LD, a process that would be consistent with our data. We find this scenario less plausible, however, as there is no reason why gene flow should have been high in the past but absent in present-day sympatric populations. In fact, early generation maize-teosinte hybrids are common in these populations today (MB Hufford, pers. observation), and genetic data support ongoing gene flow between domesticated maize and both *mexicana* and *parviglumis* in a number of sympatric populations (Hufford et al., 2013; Ellstrand et al., 2007; van Heerwaarden et al., 2011).

Remaining explanations for differential frequencies of the *Hopscotch* among teosinte populations include both genetic drift and natural selection. Drift may have played a role in the San Lorenzo *parviglumis* population. Previous studies using both SSRs and genome-wide SNP data have found evidence for a population bottleneck in the San Lorenzo population (Hufford, 2010; Pyhäjärvi et al., 2013), and the lower levels of sequence diversity in this population in the 5' UTR (Region 1) coupled with more positive values of Tajima's D are consistent with these earlier findings suggesting a bottleneck. *deviations from HWE may be consistent too if we see excess of homozygotes. do we?* Such population bottlenecks can exaggerate the effects of genetic drift through which the *Hopscotch* allele may have risen to high frequency entirely by chance. This bottleneck, however, does not explain the high frequency of the *Hopscotch* in multiple populations in the Jalisco cluster. Moreover, available information on diversity and population structure among Jaliscan populations (Hufford, 2010; Pyhäjärvi et al., 2013) is not suggestive of recent colonization or other demographic events that would predict a high frequency of the allele across populations. Finally, values of the Tajima's D statistic in the 5' UTR of *tb1* are

23

suggestive of natural selection acting upon the gene in natural population of *parviglumis*. Whereas the genome-wide average of Tajima's D in genic regions of *parviglumis* is 0.45 (Hufford et al., 2012b), the statistic is quite negative in the 5' UTR of *tb1* (Table 2). This result is consistent with repeated selective sweeps near *tb1* and a putative ecological role for the gene in *parviglumis*.

do we know the Hop genotype for sequenced lines? can we separate the sequences into hop/no hop and look for differences? it wasn't until we did this that gt1 stuff really popped out. *we should know for some of them, i will check* *I've added a few sentences on selection. Do we still want to compare sequences with and without Hopscotch? I agree its a good idea and could end up being really interesting. Perhaps something we could look at after submission and incorporate during revisions?*

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) have 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

24

1   across traits in a three-taxa allelic series at the *tb1* locus. Studer and Doebley

2   (2012) introgressed nine unique teosinte *tb1* segments (one from *Zea diploperennis*,

3   and four each from *mexicana* and *parviglumis*) into an inbred maize background

4   and investigated phenotypic effects. Phenotypes were shown to cluster by taxon,

5   indicating *tb1* may underlie morphological diversification of *Zea*. Additional

6   analysis in Studer and Doebley (2012) suggested tillering index was controlled both

7   by *tb1* and loci elsewhere in the genome. Clues to the identity of these loci may be

8   found in QTL studies that have identified loci controlling branching architecture

9   (*e.g.,* Doebley and Stec 1991, 1993). Many of these loci (*grassy tillers*, *gt1*;

10   *tassel-replaces-upper-ears1*, *tru1*; *terminal ear1*, *ter1*) have been shown to interact

11   with *tb1* (Whipple et al., 2011; Li, 2012), and both *tru1* and *ter1* affect the same

12   phenotypic traits as *tb1* (Doebley et al., 1995). *tassel-replaces-upper-ears1* (*tru1*),

13   for example, has been shown to act either epistatically or downstream of *tb1*,

14   affecting both branching architecture (decreased apical dominance) and tassel

15   phenotypes (shortened tassel and shank length and reduced tassel number; Li

16   2012). Variation in these additional loci may have affected tillering in our

17   collections and contributed to the lack of correlation we see between *Hopscotch*

18   genotype and tillering.

19   In conclusion, our findings demonstrate that the *Hopscotch* allele is more

20   widespread in populations of *parviglumis* and *mexicana* than previously thought.

21   Analysis of linkage using SNPs from across chromosome 1 does not suggest that the

22   *Hopscotch* allele is present in these populations due to recent introgression;

23   however, it seems unlikely that the insertion would have drifted to high frequency

24   in multiple populations. We do, however, find preliminary evidence of selection on

25   the *tb1* locus in *parviglumis*; this coupled with our observation of high frequency of

26   the *Hopscotch* insertion in a number of populations suggests that the locus plays an

1 ecological role in teosinte. In contrast to domesticated maize, the *Hopscotch*

2 insertion in *parviglumis* does not appear to reduce tillering. Other loci involved in

3 branching architecture may regulate tillering in teosinte. Future studies should

4 examine expression levels of *tb1* in teosinte with and without the *Hopscotch*

5 insertion and further characterize the effects of additional loci involved in

6 branching architecture (e.g. *gt1*, *tru1*, and *ter1*). These data, in conjunction with

7 more exhaustive phenotyping, should help reveal the ecological significance of the

8 domesticated *tb1* allele in natural populations of teosinte. *why not Phyb and phya?   Are*

9 *they necessary to include? I'd had them in before in a paragraph but had been voted out   I'd ditch gt1 tru1 ter1*

10 *and maybe just cite some people including phyb etc.*

11 *please check format of supp figs and tables; some are running off the page. you can use "longtable" to fix that*

12 *(ask Paul for example). check fig/table references, bibliography, etc. what does "rotation" mean in supp. table 3?*

13 *it isn't mentioned in methods. please check that all the tables and figs (including supplement) are referenced in the*

14 *text.*

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.

BAI, Y. AND P. LINDHOUT. 2007. Domestication and breeding of tomatoes: What have we gained and what can we gain in the future? *Annals of Botany* 100: 1085–1094.

BARRETT, S. 1983. Crop mimicry in weeds. *Econ Bot* 37: 255–282.

BRADBURY, P., Z. ZHANG, D. KROON, T. CASSTEVENS, 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. KAEPPLER, 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. HOSEMANN, 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.

DOEBLEY, J. 2004. The genetics of maize evolution. *Annu Rev Genet* 38: 37–59.

DOEBLEY, J. AND A. STEC. 1991. Genetic-analysis of the morphological differences between maize and teosinte. *Genetics* 129: 285–295.

DOEBLEY, J. AND A. STEC. 1993. Inheritance of the morphological differences between maize and teosinte: Comparison of results for two f$_2$ populations. *Genetics* 134: 559–570.

DOEBLEY, J., A. STEC, AND C. GUSTUS. 1995. *teosinte branched1* and the origin of maize: Evidence for epistasis and the evolution of dominance. *Genetics* 141: 333–346.

DOEBLEY, J., A. STEC, AND L. HUBBARD. 1997. The evolution of apical dominance in maize. *Nature* 386: 485–488.

1   DOYLE, J. AND J. DOYLE. 1990. A rapid total dna preparation procedure for

2   small quantities of fresh tissue. *Phytochemical Bulletin* 19: 11–15.

3   ELLSTRAND, N., L. GARNER, S. HEDGE, R. GUADAGNUOLO, AND

4   L. BLANCAS. 2007. Spontaneous hybridization between maize and teosinte.

5   *Journal of Heredity* 98: 183–187.

6   ELLSTRAND, N., H. PRENTICE, AND J. HANCOCK. 1999. Gene flow and

7   introgression from domesticated plants into their wild relatives. *Annu Rev Ecol*

8   *Syst* 30: 539–563.

9   FALUSH, D., M. STEPHENS, AND J. PRITCHARD. 2003. Inference of population

10   structure using multilocus genotype data: Linked loci and correlated allele

11   frequencies. *Genetics* 164: 1567–1587.

12   FAO/IIASA/ISRIC/ISSCAS/JRC. 2012. Harmonized World Soil Database,

13   version 1.2. FAO, Rome, Italy and IIASA, Laxenburg, Austria.

14   FEDER, J., S. BERLOCHER, J. ROETHELE, H. DAMBROSKI, J. SMITH,

15   W. PERRY, V. GAVRILOVIC, K. FILCHAK, J. RULL, AND M. ALUJA. 2003.

16   Allopatric genetic origins for sympatric host-plant shifts and race formation in

17   rhagoletis. *P Natl Acad Sci Usa* 100: 10314–10319.

18   FOX, J. AND S. WEISBERG. 2011. An R Companion to Applied Regression, vol.

19   Second Edition. Sage, Thousand Oaks, CA.

20   FUKUNAGA, N.-W. T. L. B. Z. Q. V. Y. F. M. B. K. L. L. D. J., K. 2005.

21   Genetic diversity and population structure of teosinte. *Genetics* 169: 2241–2254.

22   GALLAVOTTI, A., Q. ZHAO, J. KYOZUKA, R. MEELEY, M. RITTER,

23   J. DOEBLEY, M. PE, AND R. SCHMIDT. 2004. The role of barren stalk1 in the

24   architecture of maize. *Nature* 432: 630–635.

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

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

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

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

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

15 HUFFORD, M., P. LUBINSKY, T. PYHÄJÄRVI, M. DEVENGENZO,
16 N. ELLSTRAND, AND J. ROSS-IBARRA. 2013. The genomic signature of
17 crop-wild introgression in maize. *PLoS Genetics* 9: e1003477.

18 HUFFORD, M., X. XU, J. VAN, HEERWAARDEN, T. PYHÄJÄRVI, J. CHIA,
19 R. CARTWRIGHT, R. ELSHIRE, J. GLAUBITZ, K. GUILL, S. KAEPPLER,
20 J. LAI, P. MORRELL, L. SHANNON, C. SONG, N. SPRINGER,
21 R. SWANSON-WAGNER, P. TIFFIN, J. WANG, G. ZHANG, J. DOEBLEY,
22 M. MCMULLEN, D. WARE, E. BUCKLER, S. YANG, AND J. ROSS-IBARRA.
23 2012b. Comparative population genomics of maize domestication and
24 improvement. *Nat Genet* 44: 808–U118.

Kearse, M., R. Moir, A. Wilson, S. Stones-Havas, M. Cheung, S. Sturrock, S. Buxton, A. Cooper, S. Markowitz, C. Duran, T. Thierer, B. Ashton, P. Meintjes, and A. Drummond. 2012. Geneious basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28: 1647–1649.

Kebrom, T. and T. Brutnell. 2007. The molecular analysis of the shade avoidance syndrome in the grasses has begun. *Journal of Experimental Botany* 58: 3079–3089.

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

Kovach, M. and S. McCouch. 2008. Leveraging natural diversity: back through the bottleneck. *Genome studies and Molecular Genetics* 11: 193–200.

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

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

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

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

1   OLSEN, K. AND B. GROSS. 2010. Genetic perspectives on crop domestication.

2   *Trends in Plant Science* 15: 529–537.

3   PIPERNO, D., A. RANERE, I. HOLST, J. IRIARTE, AND R. DICKAU. 2009.

4   Starch grain and phytolith evidence for early ninth millennium bp maize from the

5   central balsas river valley, mexico. *P Natl Acad Sci Usa* 106: 5019–5024.

6   PLANTINGA, T., S. ALONSO, N. IZAGIRRE, M. HERVELLA, R. FREGEL,

7   J. VAN DER MEER, M. NETEA, AND C. DE LA RUA. 2012. Low prevalence of

8   lactase persistence in neolithic south-west europe. *Eur J Hum Genet* 20: 778–782.

9   PYHÄJÄRVI, T., M. HUFFORD, AND J. ROSS-IBARRA. 2013. Complex patterns of

10   local adaptation in the wild relatives of maize. *Genome Biology and Evolution* 5:

11   1594–1609.

12   ROSS-IBARRA, J., P. MORRELL, AND B. GAUT. 2007. Plant domestication, a

13   unique opportunity to identify the genetic basis of adaptation. *P Natl Acad Sci*

14   *Usa* 104: 8641–8648.

15   ROSS-IBARRA, J., M. TENAILLON, AND B. GAUT. 2009. Historical divergence

16   and gene flow in the genus zea. *Genetics* 181: 1399–1413.

17   ROZEN, S. AND H. SKALETSKY. 2000. Primer3 on the www for general users and

18   for biologist programmers. *Methods in Molecular Biology* : 365–386.

19   SIGMON, B. AND E. VOLLBRECHT. 2010. Evidence of selection at the ramosa1

20   locus during maize domestication. *Mol Ecol* 19: 1296–1311.

21   STUDER, A. AND J. DOEBLEY. 2012. Evidence for a natural allelic series at the

22   maize domestication locus teosinte branched1. *Genetics* 19: 951–958.

1 STUDER, A., Q. ZHAO, J. ROSS-IBARRA, AND J. DOEBLEY. 2011. Identification

2 of a functional transposon insertion in the maize domestication gene tb1. *Nat*

3 *Genet* 43: 1160–U164.

4 THORNTON, K. 2003. libsequence: a c++ class library for evolutionary genetic

5 analysis. *Bioinformatics* 19: 2325–2327.

6 THURBER, C., M. REAGON, B. GROSS, K. OLSEN, Y. JIA, AND A. CAICEDO.

7 2010. Molecular evolution of shattering loci in us weedy rice. *Mol Ecol* 19:

8 3271–3284.

9 TISHKOFF, S., F. REED, A. RANCIARO, B. VOIGHT, C. BABBITT,

10 J. SILVERMAN, K. POWELL, H. MORTENSEN, J. HIRBO, M. OSMAN,

11 M. IBRAHIM, S. OMAR, G. LEMA, T. NYAMBO, J. GHORI, S. BUMPSTEAD,

12 J. PRITCHARD, G. WRAY, AND P. DELOUKAS. 2007. Convergent adaptation of

13 human lactase persistence in africa and europe. *Nat Genet* 39: 31–40.

14 VAN HEERWAARDEN, J., J. DOEBLEY, W. BRIGGS, J. GLAUBITZ,

15 M. GOODMAN, J. GONZALEZ, AND J. ROSS-IBARRA. 2011. Genetic signals of

16 origin, spread, and introgression in a large sample of maize landraces. *P Natl*

17 *Acad Sci Usa* 108: 1088–1092.

18 WEBER, A., R. CLARK, L. VAUGHN, J. SANCHEZ-GONZALEZ, J. YU,

19 B. YANDELL, P. BRADBURY, AND J. DOEBLEY. 2007. Major regulatory genes

20 in maize contribute to standing variation in teosinte (zea mays ssp parviglumis).

21 *Genetics* 177: 2349–2359.

22 WHIPPLE, C., T. KEBROM, A. WEBER, F. YANG, D. HALL, R. MEELEY,

23 R. SCHMIDT, J. DOEBLEY, T. BRUTNELL, AND D. JACKSON. 2011. grassy

tillers1 promotes apical dominance in maize and responds to shade signals in the grasses. *P Natl Acad Sci Usa* 108: E506–E512.

WHITTON J, A. D. S. A. R. L., WOLF DE. 1997. The persistence of cultivar alleles in wild populations of sunflowers fiver generations after hybridization. *Theoretical and Applied Genetcs* 95: 33–40.

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

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

Supplementary Materials

5'　　　　　　　　　Hopscotch　　　　　　　//　　　　　　　tbl　　　3'

1,235bp　　　　　　　　　　　　1,335bp

58.5kb
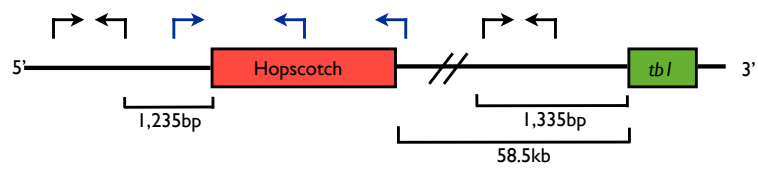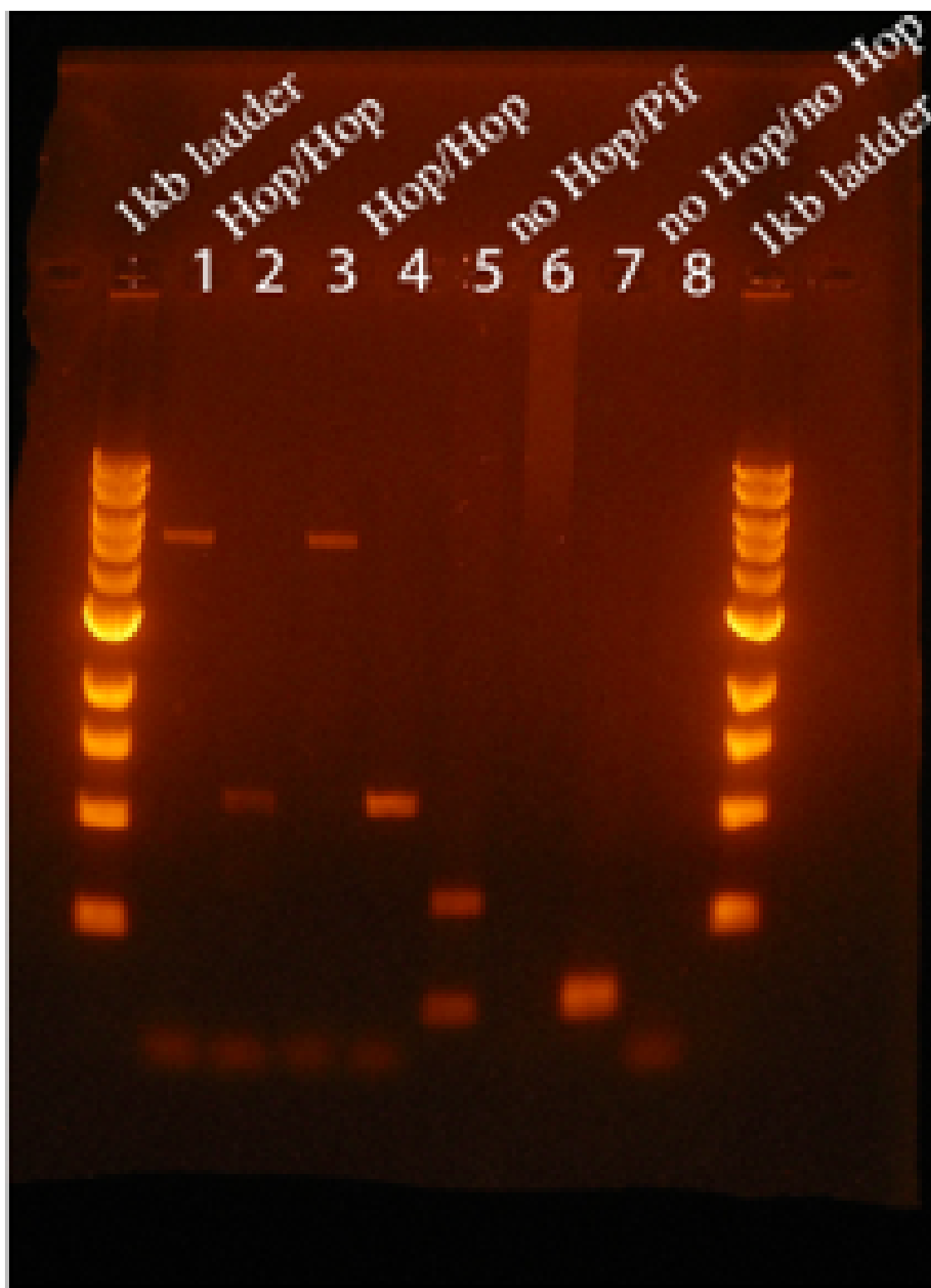
Accession	USDA Accession ID	Locality	Number alleles sampled	Hopscotch
RIHY0009	N/A	N/A	2	
RIMME0006	566673	Durango, Mexico	2	
RIMME0007	566680	Guanajuato, Mexico	2	
RIMME0008	566681	Michoacan, Mexico	2	
RIMME0009	566682	Distrito Federal, Mexico	2	
RIMME0011	566685	Mexico, Mexico	2	
RIMME0014	714151	Breeders line; Puga: 11066	6	
RIMME0017	699874	Ayotlan, Mexico	8	
RIMME0021	N/A	El Porvenir, Mexico	69	0
RIMME0026	N/A	Opopeo, Mexico	42	0
RIMME0028	N/A	Puruandiro, Mexico	28	0
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	
RIMPA0045	288193	Guerrero, Mexico	4	
RIMPA0055	714152	Breeders line	2	
RIMPA0056	714153	Breeders line	2	
RIMPA0057	714154	Breeders line	2	
RIMPA0058	N/A	N/A	4	

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	39	1
RIMMA0149	2		1
RIMMA0152	2		1
RIMMA0153	2		1
RIMMA0154	2		1
RIMMA0155	2		1

	PC1		PC2		PC3		PC4		PC5		PC
	Var	Rot	Var	Rot	Var	Rot	Var	Rot	Var	Rot	Va
bio1	0.146	bio4	0.244	prec7	0.287	ts_clay	0.41	bio2	0.38	bi
tmean11	0.146	bio3	0.241	prec8	0.276	v_mod	0.359	sq4	0.328	x_m
tmean12	0.145	bio7	0.241	prec11	0.262	ts_sand	0.329	ts_loam	0.289	s
bio11	0.145	prec6	0.237	bio13	0.247	bio15	0.272	ts_sand	0.266	bi
tmax12	0.145	sq7	0.218	prec1	0.246	prec4	0.259	sq7	0.231	v_m
tmin5	0.145	prec9	0.217	bio16	0.242	x_mod	0.244	bio18	0.213	prec
tmean1	0.145	sq3	0.207	prec12	0.24	prec3	0.226	bio13	0.207	bio
tmean2	0.145	prec12	0.207	bio19	0.238	sq3	0.21	prec11	0.183	s
tmin4	0.145	bio12	0.204	bio12	0.231	prec5	0.21	bio7	0.17	s
tmax1	0.145	bio19	0.196	prec2	0.222	prec7	0.19	bio16	0.163	ts_sar
tmean4	0.145	prec2	0.188	bio18	0.221	sq4	0.186	bio4	0.157	bi
tmin11	0.144	prec1	0.185	sq4	0.2	bio3	0.185	bio12	0.156	prec
tmax11	0.144	prec10	0.184	prec9	0.18	bio18	0.178	bio3	0.155	tmax
tmin12	0.144	bio16	0.183	prec10	0.171	sq7	0.132	prec6	0.154	tmax
tmin2	0.144	prec8	0.17	prec5	0.161	bio14	0.116	x_mod	0.152	bi
tmean5	0.144	prec5	0.165	prec4	0.154	bio13	0.099	prec9	0.144	tmax
tmean10	0.144	bio14	0.158	sq3	0.147	bio16	0.095	prec8	0.143	bi
bio6	0.144	bio13	0.151	bio2	0.143	prec8	0.09	v_mod	0.142	ts_loa
tmax2	0.144	bio17	0.149	bio17	0.129	bio7	0.077	bio15	0.136	ts_cl
tmean3	0.144	prec3	0.144	ts_loam	0.127	bio4	0.075	prec7	0.112	tmin
tmin1	0.143	ts_clay	0.141	v_mod	0.123	bio2	0.074	prec4	0.108	tmin
tmin10	0.143	bio2	0.129	prec3	0.113	prec2	0.074	bio14	0.096	prec
Altitude	0.143	prec7	0.108	x_mod	0.111	bio19	0.068	tmax7	0.093	tmin1
bio9	0.143	tmax6	0.107	bio14	0.099	prec12	0.056	tmax8	0.092	tmin1
tmin3	0.143	x_mod	0.106	bio4	0.07	ts_loam	0.053	prec1	0.091	tmin
bio10	0.142	bio15	0.098	tmax3	0.067	tmax12	0.047	prec2	0.086	tmean
tmax10	0.142	ts_loam	0.088	ts_clay	0.065	bio17	0.047	tmin11	0.086	tmax
tmax3	0.142	tmean6	0.085	bio15	0.056	bio9	0.043	prec5	0.082	tmax
tmax4	0.142	tmin7	0.082	tmax2	0.055	tmax8	0.042	bio17	0.082	tmean
tmin6	0.142	bio5	0.082	tmean3	0.052	tmax1	0.041	tmin12	0.08	bi
tmean9	0.141	tmean7	0.081	ts_sand	0.05	tmax5	0.039	prec3	0.078	tmin1
tmin9	0.141	prec4	0.08	prec6	0.048	tmax7	0.039	tmax9	0.078	prec

40

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