

1 Natural variation in teosinte at the domestication locus

2 *teosinte branched1* (*tb1*)

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

5 ¹Department of Plant Sciences, University of California Davis

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

7 ³Department of Biology, University of Oulu

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

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

10 May 16, 2014

*mhufford@iastate.edu

†rossibarra@ucdavis.edu

1 Abstract

2 The *teosinte branched1* (*tb1*) gene, a repressor of lateral organ growth, is a major QTL
3 involved in branching differences between maize and its wild progenitor, teosinte. Further
4 studies have shown that the insertion of a transposable element (*Hopscotch*) upstream of
5 *tb1* enhances its expression, causing the reduction in branching observed in domesticated
6 maize. Observations of the maize *tb1* allele in teosinte individuals, coupled with estimates
7 of the age of insertion of the *Hopscotch* element, led us to investigate the prevalence and
8 potential role of *tb1* in teosinte. Results from genotyping across many natural
9 populations suggest that the *Hopscotch* element is segregating at a higher than expected
10 frequency in a number of populations of two subspecies of teosinte, *Zea mays* ssp.
11 *parviglumis* and *Zea mays* ssp. *mexicana*. Analysis of linkage disequilibrium between the
12 *Hopscotch* element and variation in surrounding regions does not support a hypothesis of
13 recent introgression from maize into teosinte, and we find no evidence of environmental
14 correlations that might suggest recent selection. Finally, two greenhouse experiments
15 with *Zea mays* ssp. *parviglumis* do not suggest an important role for *tb1* in controlling
16 tillering in natural populations of this subspecies. Our findings suggest that the role of
17 the *Hopscotch* in tillering in teosinte is not as straightforward as is in domesticated
18 maize, and that other loci may play a role in observed variation in this trait.

INTRODUCTION

Domesticated crops and their wild progenitors provide an excellent system in which to study adaptation and genomic changes associated with human-mediated selection (Ross-Ibarra et al., 2007). 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 (Olsen and Gross, 2010). 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 (Kovach and McCouch, 2008). But while some effort has been invested into understanding how wild alleles behave in their domesticated relatives (Bai and Lindhout, 2007), very little is known about the role that alleles found most commonly in domesticates play in natural populations of their wild progenitors. (Whitton J, 1997).

Maize (*Zea mays* ssp. *mays*) was domesticated from the teosinte *Zea mays* ssp. *parviglumis* (hereafter, *parviglumis*) roughly 9,000 B.P. in southwest Mexico (Piperno et al., 2009; Matsuoka et al., 2002). 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 (Hufford et al., 2012a; Doebley, 2004; Hufford et al., 2012b). Additionally, large naturally occurring populations of both *Zea mays* ssp. *parviglumis* (the wild progenitor of maize) and *Zea mays* ssp. *mexicana* (highland teosinte; hereafter *mexicana*) can be found throughout Mexico (Wilkes, 1977; Hufford et al., 2013), and genetic diversity of these taxa is estimated to be high (Ross-Ibarra et al., 2009).

Many morphological changes are associated with maize domestication, and understanding the genetic basis of these changes has been a focus of maize research for a number of years (Doebley, 2004). One of the most dramatic changes is found in plant architecture: domesticated maize is characterized by a central stalk with few tillers and

1 lateral branches terminating in a female inflorescence, while teosinte is highly tillered and
2 bears tassels (male inflorescences) at the end of its lateral branches. The *teosinte*
3 *branched1* (*tb1*) gene, a repressor of organ growth, was identified as a major QTL
4 involved in domestication branching differences achieved through higher expression of the
5 maize allele than the teosinte allele (Doebley et al., 1995, 1997). Further work showed
6 that the insertion of a 4.9 kb retrotransposon (*Hopscotch*) in the upstream control region
7 of *tb1* caused the increased expression of this gene reduction in branching observed in
8 domesticated maize Studer et al. (2011). The effects of this insertion have been observed
9 in tiller number in maize, but little is known about its role, if any, in natural populations
10 of teosinte (Studer et al., 2011). Dating of this element has suggested that its insertion
11 predates the domestication of maize, leading to the hypothesis that it was segregating as
12 standing variation in ancient populations of teosinte and increased to high frequency in
13 maize due to selection during domestication (Studer et al., 2011). Furthermore, Studer
14 and Doebley (2012) investigated the phenotypic effects of nine teosinte *tb1* alleles in an
15 isogenic maize background and found that the introgressions sort into three distinct
16 phenotypic classes, suggesting that variation at the *tb1* locus may play a functional role
17 in teosinte.

18 In high-density species such as teosinte, plants can detect impending competition
19 from their neighbors through detection of the ratio of red to far-red light. An increase in
20 far-red relative to red light accompanies shading and triggers physiological and
21 morphological changes such as reduced tillering, increased plant height and early
22 flowering collectively known as the shade avoidance syndrome (Kebrom and Brutnell,
23 2007). The *tb1* locus appears to play an important role in the shade avoidance pathway
24 in *Zea mays* and other grasses and may therefore be crucial to the ecology of teosinte
25 (Kebrom and Brutnell, 2007; Lukens and Doebley, 1999). In this study we aim to
26 characterize the distribution of the *Hopscotch* insertion in *parviglumis*, *mexicana*, and
27 landrace maize, and to examine the phenotypic effects of the insertion in *parviglumis*. We

1 use a combination of PCR genotyping for the *Hopscotch* element in our full panel and
2 sequencing of two small regions upstream of *tb1* in a subset of teosinte populations to
3 explore patterns of genetic variation at this locus. Finally, we test for an association
4 between the *Hopscotch* element and tillering phenotypes in a population of *parviglumis*.

5 MATERIALS AND METHODS

6 *from AJB formatting instructions: add name, city, spelled-out state (if in USA), and country of*
7 *manufacturers/suppliers after brand names*

8 Sampling and genotyping

9 We sampled 1,110 individuals from 350 accessions (247 maize landraces, 17 *mexicana*
10 populations, and 86 *parviglumis* populations) and assessed the presence or absence of the
11 *Hopscotch* insertion (Table S1 and Table S2). DNA was extracted from leaf tissue using a
12 modified CTAB approach (Doyle and Doyle, 1990; Maloof et al., 1984). We designed
13 primers using PRIMER3 (Rozen and Skaletsky, 2000) implemented in Geneious (Kearse
14 et al., 2012) to amplify the entire *Hopscotch* element, as well as an internal primer
15 allowing us to simultaneously check for possible PCR bias between presence and absence
16 of the *Hopscotch* insertion. Two PCRs were performed for each individual, one with
17 primers flanking the *Hopscotch* (HopF/HopR) and one with a flanking primer and an
18 internal primer (HopF/HopIntR). Primer sequences are HopF,
19 5'-TCGTTGATGCTTTGATGGATGG-3'; Hop R, 5'-AACAGTATGATTTTCATGGGACCG-3';
20 and HopIntR, 5'-CCTCCACCTCTCATGAGATCC-3' (Fig. S1, Fig. S2) *Primers in Fig. S1 should*
21 *be labeled* . Homozygotes show a single band for absence of the element (~300bp) and two
22 bands for presence of the element (~5kb and XX *LV, please add the size of the second band*),
23 whereas heterozygotes are three-banded (Fig. S2). When only one PCR resolved well, we
24 scored one allele for the individual. We used Phusion High Fidelity Enzyme (Finnzymes,
25 Inc.) and the following conditions for amplifications: 98°C for 3 min, 30 cycles of 98°C

1 for 15 s, 65°C for 30 s, and 72°C for 3 min 30 s, with a final extension of 72°C for 10 min.
2 PCR products were visualized on a 1% agarose gel and scored for presence/absence of the
3 *Hopscotch* based on band size.

4 Sequencing

5 In addition to genotyping, we chose a subset of *parviglumis* individuals for sequencing.
6 We chose twelve individuals from each of four populations from Jalisco state, Mexico
7 (San Lorenzo, La Mesa, Ejutla A, and Ejutla B). For amplification and sequencing, we
8 selected two regions approximately 600bp in size from within the 5' UTR of *tb1* (Region
9 1) and from 1,235bp upstream of the start of the *Hopscotch* (66,169bp upstream from the
10 start of the *tb1* ORF; Region 2). We designed the following primers using PRIMER3
11 (Rozen and Skaletsky, 2000): for the 5' UTR, 5'-GGATAATGTGCACCAGGTGT-3' and
12 5'-GCGTGCTAGAGACACYTGTTGCT-3'; for the 50kb upstream region,
13 5'-TGTCCTCGCCGCAACTC-3' and 5'-TGTACGCCCCGCCCTCATCA-3' (Fig. S1). We
14 used Taq polymerase (New England Biolabs) and the following thermal cycler conditions
15 to amplify fragments: 94°C for 3 min, 30 cycles of 92°C for 40 s, annealing for 1 min,
16 72°C for 40 s, and a final 10 min extension at 72°C. Annealing temperatures for
17 sequenced region 1 and sequenced region 2 were 59.7°C and 58.8°C, respectively. To
18 clean excess primer and dNTPs we added two units of Exonuclease1 and 2.5 units of
19 Antarctic Phosphatase to 8.0 μ L of amplification product. This mix was placed on a
20 thermal cycler with the following program: 37°C for 30 min, 80°C for 15 min, and a final
21 cool-down step to 4°C.

22 We cloned cleaned fragments into a TOPO-TA vector (Invitrogen, Carlsbad) using
23 OneShot TOP10 chemically competent *E. coli* cells, with an extended ligation time of 30
24 min for a complex target fragment. We plated cells on LB agar plates containing
25 kanamycin, and screened colonies using vector primers M13 Forward and M13 Reverse
26 under the following conditions: 96°C for 5 min; then 35 cycles at 96°C for 30 s, 53°C for

1 30 s, 72°C for 2 min; and a final extension at 72°C for 4 min. We visualized amplification
2 products for incorporation of our insert on a 1% agarose TAE gel.

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

14 Genotyping analysis

15 We examined discrepancies between observed and expected genotype frequencies by
16 calculating Hardy-Weinberg Equilibrium (HWE). To calculate differentiation between
17 populations (F_{ST}) and subspecies (F_{CT}) we used HierFstat (Goudet, 2005). These
18 analyses only included populations in which 8 or more individuals were sampled. To test
19 the hypothesis that the *Hopscotch* insertion may be adaptive under certain environmental
20 conditions, we looked for significant associations between the *Hopscotch* frequency and
21 environmental variables using BayEnv (Coop et al., 2010). BayEnv creates a covariance
22 matrix of relatedness between populations and then tests a null model that allele
23 frequencies in populations are determined by the covariance matrix of relatedness alone
24 against the alternative model that allele frequencies are determined by a combination of
25 the covariance matrix and an environmental variable, producing a posterior probability
26 (*i.e.*, Bayes Factor; Coop et al. 2010). We used genotyping and covariance data from

1 Pyhäjärvi et al. (2013) for BayEnv, with the *Hopscotch* insertion coded as an additional
2 SNP (Table S3). Environmental data were obtained from www.worldclim.org, the
3 Harmonized World Soil Database (FAO/IIASA/ISRIC/ISSCAS/JRC, 2012) and
4 www.harvestchoice.org and summarized by principle component analysis (Pyhäjärvi
5 et al., 2013).

6 Sequence analysis

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

19 Introgression analysis

20 In order to assess patterns of linkage disequilibrium (LD) around the *Hopscotch* element
21 in the context of chromosomal patterns of LD we used Tassel (Bradbury et al., 2007) and
22 calculated LD between SNPs across chromosome 1 using previously published data from
23 twelve plants each of the Ejutla A (EjuA), Ejutla B (EjuB), San Lorenzo (SLO), and La
24 Mesa (MSA) populations (Pyhäjärvi et al., 2013). We chose these populations because we
25 had both genotyping data for the *Hopscotch* as well as chromosome-wide SNP data for

1 chromosome 1. For each population we filtered the initial set of 5,897 SNPs on
2 chromosome 1 to accept only SNPs with a minor allele frequency of at least 0.1, resulting
3 in 1,671, 3,023, 3,122, and 2,167 SNPs for SLO, EjuB, EjuA, and MSA, respectively. We
4 then used Tassel (Bradbury et al., 2007) to calculate linkage disequilibrium (r^2) across
5 chromosome 1 for each population.

6 We examined evidence of introgression on chromosome 1 in these same four
7 populations (EjuA, EjuB, MSA, SLO) using STRUCTURE (Falush et al., 2003) and the
8 same phased 55K SNP data from (Pyhäjärvi et al., 2013) that we used for LD analysis,
9 combined with the corresponding SNP data from a diverse panel of 282 maize lines (Cook
10 et al., 2012). SNPs were anchored in a modified version of the IBM genetic map (Gerke
11 et al., 2013). We created haplotype blocks using a custom Perl script that grouped SNPs
12 separated by less than 5kb into haplotypes. We ran STRUCTURE at K=2 under the
13 linkage model, performing 3 replicates with an MCMC burn-in of 10,000 steps and 50,000
14 steps post burn-in.

15 Phenotyping of *parviglumis*

16 To investigate the phenotypic effects of the *Hopscotch* insertion in teosinte, we conducted
17 an initial phenotyping trial (Phenotyping 1). We germinated 250 seeds of *parviglumis*
18 collected in Jalisco state, Mexico (population San Lorenzo) (Hufford, 2010) where the
19 *Hopscotch* is segregating at highest frequency (0.44) in our initial genotyping sample set.
20 In order to maximize the likelihood of finding the *Hopscotch* in our association
21 population we selected seeds from sites where genotyped individuals were homozygous or
22 heterozygous for the insertion. We chose between 10-13 seeds from each of 23 sampling
23 sites. We treated seeds with fungicide and germinated them in petri dishes with filter
24 paper. Following germination, 206 successful germinations were then planted into
25 one-gallon pots with potting soil and randomly spaced one foot apart on greenhouse
26 benches. Plants were watered three times a day by hand and with an automatic drip

1 containing 10-20-10 fertilizer.

2 Starting on day 15, we measured tillering index as the ratio of the sum of tiller
3 lengths to the height of the plant (Briggs et al., 2007). Following initial measurements,
4 we phenotyped plants for tillering index every 5 days through day 40, and then on day 50
5 and day 60. On day 65 we measured culm diameter between the third and fourth nodes
6 of each plant. Culm diameter is not believed to be correlated with tillering index or
7 variation at *tb1*. Following phenotyping we extracted DNA from all plants using a
8 modified SDS extraction protocol (<http://www.ars.usda.gov>). We genotyped
9 individuals for the *Hopscotch* insertion following the protocols listed above. Based on
10 these initial data, we conducted a post hoc power analysis using data from day 40 of
11 Phenotyping 1, indicating that a minimum of 71 individuals in each genotypic class are
12 needed to detect the observed effect of the *Hopscotch* on tillering index.

13 We performed a second phenotyping experiment (Phenotyping 2) in which we
14 germinated 372 seeds of *parviglumis*, choosing equally between sites previously
15 determined to have or not have the *Hopscotch* insertion. Seeds were germinated and
16 planted on day 7 post fruit-case removal into two gallon pots. Plants were watered twice
17 daily, alternating between fertilized and non-fertilized water. We began phenotyping
18 successful germinations (302) for tillering index on day 15 post fruit-case removal, and
19 phenotyped every five days until day 50. At day 50 we measured culm diameter between
20 the third and fourth nodes. We extracted DNA and genotyped plants following the same
21 guidelines as in Phenotyping 1.

22 Resulting tillering index data for each genotypic class did not meet the criteria for a
23 repeated measures ANOVA, so we transformed the data using a Box-Cox transformation
24 ($\alpha = 0$) *what is the alpha value here?* implemented in the car package in R (Fox and Weisberg,
25 2011) to improve the normality and homogeneity of variance among genotype classes. We
26 analyzed relationships between genotype and tillering index and tiller number using a
27 repeated measures ANOVA through a general linear model function implemented in SAS

1 v.9.3 (SAS Institute Inc., Cary, NC, USA). Additionally, in order to compare any
2 association between *Hopscotch* genotype and tillering and associations at other
3 presumably unrelated traits, we performed an ANOVA between culm diameter and
4 genotype using the same general linear model in SAS.

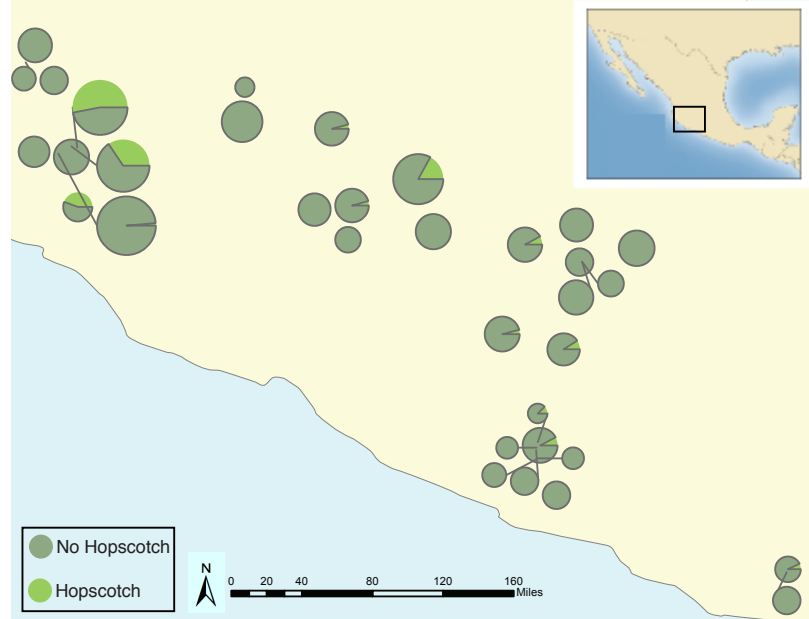
5 RESULTS

6 Genotyping

7 Genotype of the *Hopscotch* insertion was confirmed with two PCRs for 837 individuals.
8 Among the 247 maize landrace accessions genotyped, all but eight were homozygous for
9 the presence of the insertion (Table S1 and Table S2). Within our *parviglumis* and
10 *mexicana* samples we found the *Hopscotch* insertion segregating in 37 and 4 populations,
11 respectively, and at highest frequency in the states of Jalisco, Colima, and Michoacán in
12 central-western Mexico (Fig.1). We examined Hardy-Weinberg equilibrium in a total of
13 14 populations (10 *parviglumis* and 4 *mexicana*) with more than 8 individuals sampled
14 per population. Three populations (RIMPA0073, RIMPA0093, and RIMPA0158) show
15 evidence of deviations from expected genotype frequencies under the assumptions of
16 HWE ($p < 0.05$). *Jeff mentioned this needs to be fully reported in a supplementary table; alternatively we can*
17 *just delete the result since we do not interpret or use it later in the manuscript.*

18 Using our *Hopscotch* genotyping, we calculated differentiation between populations
19 (F_{ST}) and subspecies (F_{CT}) for populations in which we sampled 8 or more alleles. We
20 found that $F_{CT} = 0$ within our dataset and levels of F_{ST} among populations within each
21 subspecies (0.22) *is this an average? this comes from libsequence?* and among all populations
22 (0.23) are similar to those reported genome-wide in previous studies (Pyhäjärvi et al.
23 2013; Table 1). Although we found large variation in *Hopscotch* allele frequency among
24 our populations, BayEnv analysis did not indicate a correlation between the *Hopscotch*
25 insertion and environmental variables (all Bayes Factors < 1 ; Table S3).

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 Sequencing

2 To investigate patterns of sequence diversity and linkage disequilibrium (LD) in the *tb1*
3 region, we sequenced two small (<1kb) regions upstream of the *tb1* ORF in four
4 populations. After alignment and singleton checking we recovered 48 and 40 segregating
5 sites for the 5' UTR region (Region 1) and the 66kb upstream region (Region 2),
6 respectively. For Region 1, Ejutla A has the highest values of haplotype diversity, and θ_π ,
7 while Ejutla B and La Mesa have comparable values of these summary statistics, and San
8 Lorenzo has much lower values. Additionally, Tajima's D is strongly negative in the two
9 Ejutla populations and La Mesa, but is less negative in San Lorenzo (Table 2). *need to*
10 *reference Hopscotch frequencies in supplemental table somewhere* For Region 2, haplotype diversity and
11 θ_π , are similar for Ejutla A and Ejutla B, while La Mesa and San Lorenzo have slightly
12 lower values for these statistics (Table 2). Tajima's D is positive in all populations except
13 San Lorenzo, *is the table wrong? MSA is the only negative value in the table* indicating an excess of
14 low frequency variants in this population (Table 2). Pairwise values of F_{ST} within
15 population pairs Ejutla A/Ejutla B and San Lorenzo/La Mesa are 0 for both sequenced

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

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

regions as well as for the *Hopscotch*, while they are high for other population pairs (Table 1). Neighbor joining trees of our sequence data and data from the teosinte inbred lines (TILs; data from Maize HapMapV2, Chia et al. 2012) do not reveal any clear clustering pattern with respect to population or *Hopscotch* genotype (Figure S3); 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.

7 Evidence of introgression

Teosinte populations with the highest frequency of the *Hopscotch* insertion in this study were *parviglumis* sympatric with cultivated maize. Our initial hypothesis was that the high frequency of the *Hopscotch* element in these populations could be attributed to introgression from maize into teosinte. To investigate this possibility we examined overall patterns of linkage disequilibrium across chromosome one, and specifically in the *tb1* region. If the *Hopscotch* is found in these populations due to recent introgression we would expect to find large blocks of linked markers near this element. We find no evidence of elevated linkage disequilibrium between the *Hopscotch* and SNPs surrounding the *tb1* region in our resequenced populations (2), and r^2 in the *tb1* region does not differ

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

| Population | # Haplotypes | Hap. Diversity | $\hat{\theta}_\pi$ | Tajima's D |
|---------------------------------|--------------|----------------|--------------------|------------|
| <i>Region 1 (5' UTR)</i> | | | | |
| EJUA | 8 | 0.85897 | 0.00527 | -1.64955 |
| EJUB | 5 | 0.70909 | 0.00378 | -1.83123 |
| MSA | 6 | 0.68182 | 0.00373 | -1.75506 |
| SLO | 3 | 0.31818 | 0.00137 | -0.72873 |
| <i>Region 2 (66kb upstream)</i> | | | | |
| EJUA | 8 | 0.89394 | 0.01763 | 0.6231 |
| EJUB | 8 | 0.89394 | 0.01591 | 0.29504 |
| MSA | 3 | 0.68182 | 0.01055 | -0.22212 |
| SLO | 4 | 0.74242 | 0.01413 | 0.93185 |

1 significantly between populations with (average r^2 of 0.085) and without the *Hopscotch*
2 genotype (average $r^2 = 0.082$). In fact, average r^2 is lower in the *tb1* region ($r^2 = 0.056$)
3 than across the rest of chromosome 1 ($r^2 = 0.083$) (3). *table is too wide, need to round numbers,*
4 *and column headers are messed up.*

5 The lack of clustering of *Hopscotch* genotypes in our NJ tree as well as the lack of LD
6 around *tb1* does not support the hypothesis that the *Hopscotch* insertion in these
7 populations of *parviglumis* is the result of recent introgression. However, to further
8 explore this hypothesis we performed a STRUCTURE analysis using Illumina
9 MaizeSNP50 data from four of our *parviglumis* populations (EjuA, EjuB, MSA, and
10 SLO) and the maize 282 diversity panel (Cook et al., 2012; Pyhäjärvi et al., 2013). The
11 linkage model implemented in STRUCTURE can be used to identify ancestry of blocks of
12 linked variants, which would arise as a result of recent admixture between populations. If
13 the *Hopscotch* insertion is present in populations of *parviglumis* as a result of recent

Table 3: r^2 values between SNPs in the *tb1* region (positions 264,596,664-265,891,456 on chromosome 1 of the maize AGPv2 genome) and the rest of chromosome 1, within the 5' UTR (Sequenced region 1), and within the 66,169 bp upstream region (Sequenced region 2).

| Population | Chromosome 1 | <i>tb1</i> region | Seq. region 1 | Seq. region 2 |
|-------------|--------------|-------------------|---------------|---------------|
| Ejutla A | 0.095426101 | 0.050304 | 0.747295 | 0.214933 |
| Ejutla B | 0.068681837 | 0.051295 | 0.660354 | 0.186395 |
| La Mesa | 0.069500533 | 0.053306 | 0.914286 | 0.766234 |
| San Lorenzo | 0.100536784 | 0.067251 | 0.912281 | 0.636364 |

1 admixture with domesticated maize, we would expect the insertion and linked variants in
2 surrounding sites to be assigned to the "maize" cluster in our STRUCTURE runs, not
3 the "teosinte" cluster. In all runs, assignment to maize in the *tb1* region across all four
4 *parviglumis* populations is low (average 0.017) and much below the chromosome-wide
5 average (0.20; Fig.3).

6 Phenotyping

7 To assess the contribution of *tb1* to phenotypic variation in tillering in a natural
8 population, we grew plants from seed sampled from the San Lorenzo population of
9 *parviglumis*, which had a high mean frequency (0.44) of the *Hopscotch* insertion from our
10 initial genotyping. We measured tillering index (TI), the ratio of the sum of tiller lengths
11 to plant height, for 216 plants from within the San Lorenzo population, and genotyped
12 plants for the *Hopscotch* insertion. We found the *Hopscotch* segregating at a frequency of
13 0.65 with no significant deviations from expected frequencies under Hardy-Weinberg
14 equilibrium. After performing a repeated measures ANOVA between our transformed
15 tillering index data and *Hopscotch* genotype we find a weak positive correlation between

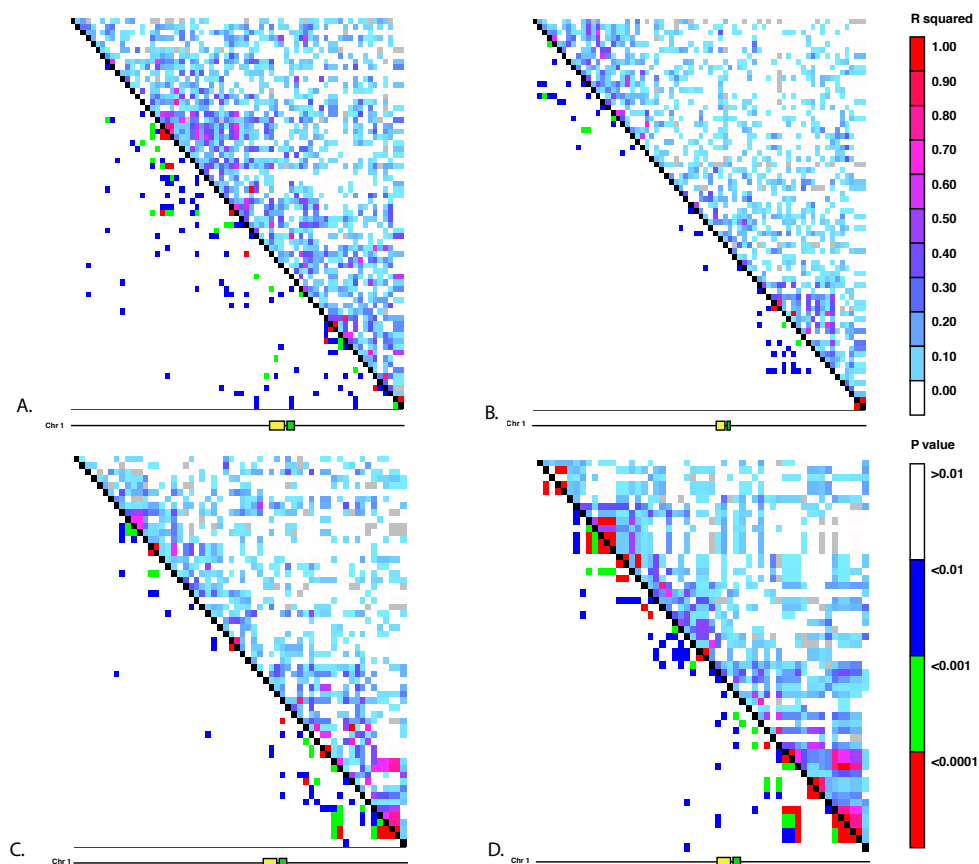


Figure 2: Linkage disequilibrium for SNPs in Mb 261-268 on chromosome 1. The yellow rectangle indicates the location of the *Hopscotch* insertion and the green represents the *tb1* ORF. A) Ejutla A; B) Ejutla B; C) La Mesa; D). San Lorenzo

- 1 presence of the *Hopscotch* and tillering index on day 40 ($p=0.0848$), but no correlation
- 2 between tillering index and genotype on any other day (4). Additionally we find no
- 3 significant correlation between tiller number and *Hopscotch* genotype, or culm diameter
- 4 and *Hopscotch* genotype in Phenotyping 1. *shouldn't we expect a negative correlation between Hop and*
- 5 *TI on day 40? need to have an A and B in the figure and explain one is for pheno1 and one is for pheno2. please*
- 6 *explain whiskers and dots on figure too. sure, I mean, presumably we would expect things with Hop to have a*

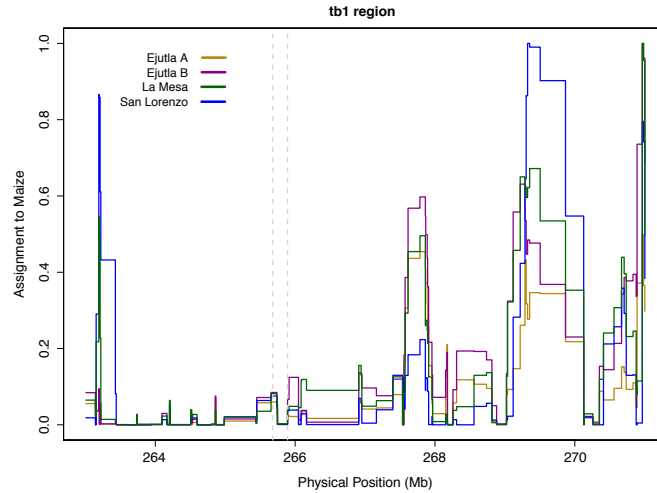


Figure 3: STRUCTURE assignment to maize across a section of chromosome 1. The dotted lines mark the beginning of the sequenced region 50kb upstream (Sequenced region 2) and the end of the *tb1* ORF.

1 *smaller TI* yup, but we should mention that the expectation is negative/
2 *lots of white space in fig 4 and fig. s1 too.* We performed a second grow-out of teosinte to
3 assess whether lighting conditions or sample size may have affected our ability to detect
4 and effect of *tb1*. For the second grow-out we measured tillering index every five days
5 through day 50 for 302 plants. We found the *Hopscotch* allele segregating at a frequency
6 of 0.69, *is it in HWE in this pop?* with a 0.6 frequency of *Hopscotch* homozygotes, and a 0.2
7 frequency of both heterozygotes and homozygotes for the teosinte allele. We found
8 similar patterns, with a weak positive correlation between tillering index and *Hopscotch*
9 genotype at day 40 (p=0.0611), with no significant correlation on any day. Similarly,
10 relationships between *Hopscotch* genotype and tiller number, and *Hopscotch* genotype
11 and culm diameter are not significant.

DISCUSSION

Adaptation occurs either due to selection on standing variation or on *de novo* mutations. Adaptation as a result of selection on standing variation has been well-described in a number of systems, for example, selection for lactose tolerance in humans (Plantinga et al., 2012; Tishkoff et al., 2007); variation at the *Eda* locus in three-spined stickleback (Kitano et al., 2008; Colosimo et al., 2005); and pupal diapause in the Apple Maggot fly (Feder et al., 2003). Although the 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*, (Gallavotti et al., 2004); and *grassy tillers1*, (Whipple et al., 2011)) 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 (Clark et al., 2006; Studer et al., 2011). Studer et al. (2011) genotyped 90 accessions of teosinte (inbred and outbred), providing the first evidence that the *Hopscotch* insertion is segregating in teosinte (Studer et al., 2011).

Given that the *Hopscotch* insertion has been estimated to predate the domestication of maize, it is not surprising that it can be found segregating in populations of teosinte. However, in sampling numerous individuals from many teosinte populations our study provides greater insight into the distribution and prevalence of the *Hopscotch* in teosinte. While our findings are consistent with a previous study by Studer et al. (2011) in that we identified the *Hopscotch* allele segregating in teosinte, we find it at higher frequency than previously suggested (Studer et al., 2011). Many of our populations with high frequency

1 of the *Hopscotch* allele fall in the Jalisco cluster identified by Fukunaga (???), possibly
2 suggesting a different history of the *tb1* locus than in the Balsas region where maize was
3 domesticated (Matsuoka et al., 2002). While gene flow from crops into their wild relatives
4 is well-known, ((Ellstrand et al., 1999; Zhang et al., 2009; Thurber et al., 2010; Baack
5 et al., 2008; Hubner et al., 2012; Wilkes, 1977; van Heerwaarden et al., 2011; Barrett,
6 1983)), our results are more consistent with Hufford et al. (2013) who found resistance to
7 introgression from maize into teosinte (Hufford et al., 2013). Furthermore, Hufford et al.
8 (2013) showed that domestication loci, such as *tb1*, are particularly resistant to
9 introgression in both directions of gene flow (i.e., maize to teosinte and teosinte to maize)
10 (Hufford et al., 2013).

11 We find no evidence of recent introgression in our analyses. Clustering patterns in
12 our NJ trees do not reflect a pattern expected if maize alleles at the *tb1* locus had
13 introgressed into populations of teosinte. Moreover, analysis of linkage in the *tb1* region
14 does not reveal patterns of high LD relative to the rest of chromosome 1, and assignment
15 to maize in this region in our STRUCTURE analysis is lower than the average across
16 chromosome 1 (3, 4). Together, these data point to an explanation other than recent
17 introgression for the high observed frequency of *Hopscotch* in some of our *parviglumis*
18 populations.

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

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

1 Although recent introgression seems unlikely, we cannot rule out ancient introgression
2 as an explanation for the presence of the *Hopscotch* in these populations. If the *Hopscotch*
3 allele was introgressed in the distant past, they could have been sufficient recombination
4 to break up any initial LD, leading to observations similar to those obtained here. We
5 find this scenario less plausible, however, as there is no reason why gene flow should have
6 been high in the past but absent in present-day sympatric populations. In fact, early
7 generation maize-teosinte hybrids are easy to find in these populations today (MB
8 Hufford, pers. observation), and genetic data support ongoing gene flow between
9 domesticated maize and both *Zea mays* ssp. *mexicana* and *Zea mays* ssp. *parviglumis* in
10 a number of sympatric populations (Hufford et al., 2013; Ellstrand et al., 2007; ?).

11 Other explanation for differential frequencies of the *Hopscotch* among teosinte
12 populations include both drift and natural selection. Previous studies using both SSRs
13 and genome-wide SNP data have found evidence for a population bottleneck in the San
14 Lorenzo population (Hufford, 2010; Pyhäjärvi et al., 2013), and the lower levels of
15 sequence diversity in the 5' UTR region and the more positive values of Tajima's D we
16 present here are consistent with these findings. *deviations from HWE may be consistent too if we see*
17 *excess of homozygotes. do we? .* This bottleneck, however, does not explain differences in
18 *Hopscotch* allele frequency among populations, and the available information on diversity
19 and population structure among these populations (Hufford, 2010; Pyhäjärvi et al., 2013)
20 is not suggestive of colonization or other demographic events that might predict a high
21 frequency of the allele in multiple populations. *here we need a few sentences on selection. the 5'*
22 *UTR has much more negative D than the upstream. do we know the Hop genotype for sequenced lines? can we*
23 *separate the sequences into hop/no hop and look for differences? it wasn't until we did this that gt1 stuff really*
24 *popped out. we should know for some of them, i will check*

25 The phenotypic effects of the *Hopscotch* insertion in domesticated maize have been
26 well documented (Clark et al., 2006; Studer et al., 2011), and Weber et al. (2007) have
27 described its effects in partially inbred lines of teosinte (Weber et al., 2007) *i don't think these*

1 *were inbred, please doublecheck.* . Our study is the first to explicitly examine the phenotypic
2 effects of the *Hopscotch* insertion in individuals sampled from a natural population of
3 teosinte. *isn't this what weber did?? for 70+ populations!?* However, we found no significant effect
4 of the *Hopscotch* on tillering index or tiller number in our phenotyping experiments, and
5 the effect of the *Hopscotch* insertion in teosinte is discordant with that of maize. The lack
6 of correlation between *Hopscotch* genotype and tillering index or tiller number is
7 surprising given its effects in maize. It is certainly possible that even though previous
8 data demonstrate an effect of the *Hopscotch* on tillering in maize (Studer et al., 2011),
9 that the effect of the *Hopscotch* in teosinte is more complicated and may be more difficult
10 to observe. Moreover, *tb1* is a single gene in a complex pathway that affects branching
11 and tillering traits, and perhaps in combination with alleles at other loci the phenotypic
12 effects of the *Hopscotch* on tillering may not be consistent. *this section still needs work. i think we*
13 *have to do more here. weber shows an association between SNPs in tb1 and branch length. we need to discuss*
14 *that!*

15 *MBH todo* Variation at *tb1* has also been shown to contribute to phenotypes other
16 than tillering (Clark et al., 2006), and a recent study by Studer and Doebley (2012)
17 examined the possibility of an allelic series at the *tb1* locus in teosinte. Studer and
18 Doebley (2012) introgressed 9 separate teosinte segments (one from *Zea diploperennis*,
19 and four from both *Zea mays* ssp. *mexicana* and *Zea mays* ssp. *parviglumis*) spanning
20 the *tb1* locus into an isogenic maize background and investigated their effects on
21 previously associated phenotypes. They found that plants with teosinte chromosomal
22 segments had greater tillering than their maize isogenic lines, and that different
23 chromosomal segments of *tb1* confer different amounts of tillering, suggesting that there
24 are multiple genetic factors in this region that affect tillering. However, in addition to
25 elucidating variance in tillering among *tb1* teosinte segments, Studer and Doebley (2012)
26 found significant variance among W22 control lines, suggesting that there are other
27 genetic factors aside from alleles at the *tb1* locus that affect tillering in maize. Doebley

1 and Stec (1991, 1993) first attempted to map QTL controlling many of the phenotypic
2 differences between domesticated maize and teosinte, and demonstrated the existence of
3 numerous QTL that contribute to the differences in branching architecture between the
4 two. Many of these loci (*grassy tillers*, *gt1*; *tassel-replaces-upper-ears1*, *tru1*; *terminal*
5 *ear1*, *ter1*) have been shown to interact with *tb1* (Whipple et al., 2011; Li, 2012), and
6 both *tru1* and *ter1* have been shown to affect the same phenotypic traits as *tb1* (Doebley
7 et al., 1995). *tassel-replaces-upper-ears1* (*tru1*), for example, has been shown to act
8 either epistatically or downstream of *tb1*, affecting both branching architecture (decreased
9 apical dominance) and tassel phenotypes (shortened tassel and shank length and reduced
10 tassel number) (Li, 2012). It seems plausible that variation in some of these other loci
11 could have affected tillering in our greenhouse population, and contributed to the lack of
12 correlation we see between *Hopscotch* genotype and tillering.

13 In summary, our findings demonstrate that the *Hopscotch* allele is more widespread
14 in populations of *parviglumis* and *mexicana* than previously thought. Analysis of linkage
15 using SNPs from across chromosome 1 does not suggest that the *Hopscotch* allele is
16 present in these populations due to recent introgression; however, it seems unlikely that it
17 would have drifted to high frequency in multiple populations and there may be another
18 explanation for the high frequency we observe in some of our populations. The *Hopscotch*
19 does not appear to have a strong effect reducing tillering in teosinte as it does in maize,
20 and other loci involved in branching architecture may also play roles in the
21 regulation of tillering in teosinte. Finally, although we see no clear evidence of recent
22 strong selection, the high frequency of the *Hopscotch* insertion in a number of
23 populations continues to suggest to us that it plays an ecological role in teosinte. In the
24 future, additional experiments will be needed to examine expression levels of *tb1* and
25 additional loci involved in branching architecture (e.g. *gt1*, *tru1*, and *ter1*) in conjunction
26 with a more exhaustive phenotyping and genotyping assay. *why not Phyb and phya? Are they*
27 *necessary to include? I'd had them in before in a paragraph but had been voted out I'd ditch gt1 tru1 ter1 and*

- 1 *maybe just cite some people including phyb etc.*
- 2 *please check format of supp figs and tables; some are running off the page. you can use "longtable" to fix that*
- 3 *(ask Paul for example). check fig/table references, bibliography, etc. what does "rotation" mean in supp. table 3?*
- 4 *it isn't mentioned in methods. please check that all the tables and figs (including supplement) are referenced in the*
- 5 *text.*

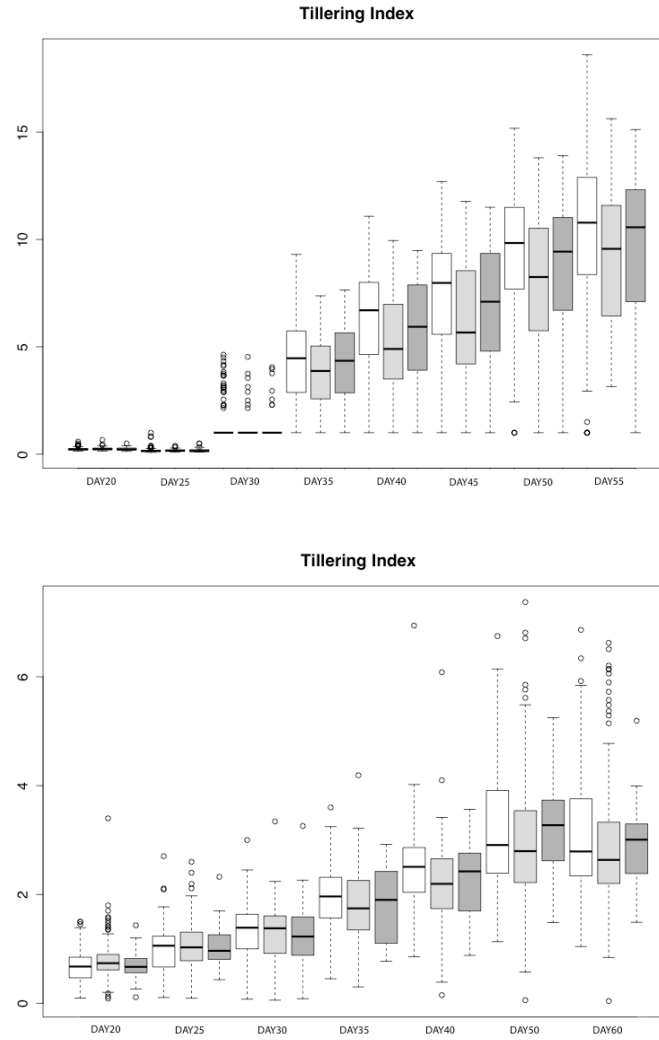


Figure 4: Box-plot showing tillering index in our greenhouse population from day 20-60. White indicates individuals homozygous for the *Hopscotch*, light grey represents heterozygotes, and dark grey represents homozygotes for the teosinte (No *Hopscotch*) allele. Within boxes, dark black lines represent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

1 References

- 2 BAACK, E., Y. SAPIR, M. CHAPMAN, J. BURKE, AND L. RIESEBERG. 2008. Selection
3 on domestication traits and quantitative trait loci in crop-wild sunflower hybrids. *Mol*
4 *Ecol* 17: 666–677.
- 5 BAI, Y. AND P. LINDHOUT. 2007. Domestication and breeding of tomatoes: What have
6 we gained and what can we gain in the future? *Annals of Botany* 100: 1085–1094.
- 7 BARRETT, S. 1983. Crop mimicry in weeds. *Econ Bot* 37: 255–282.
- 8 BRADBURY, P., Z. ZHANG, D. KROON, T. CASSTEVENS, Y. RAMDOSS, AND
9 E. BUCKLER. 2007. Tassel: software for association mapping of complex traits in
10 diverse samples. *Bioinformatics* 23: 2633–2635.
- 11 BRIGGS, W., M. McMULLEN, B. GAUT, AND J. DOEBLEY. 2007. Linkage mapping of
12 domestication loci in a large maize-teosinte backcross resource. *Genetics* 177:
13 1915–1928.
- 14 CHIA, J., C. SONG, P. BRADBURY, D. COSTICH, N. DE, LEON, J. DOEBLEY,
15 R. ELSHIRE, B. GAUT, L. GELLER, J. GLAUBITZ, M. GORE, K. GUILL,
16 J. HOLLAND, M. HUFFORD, J. LAI, M. LI, X. LIU, Y. LU, R. McCOMBIE,
17 R. NELSON, J. POLAND, B. PRASANNA, T. PYHÄJÄRVI, T. RONG, R. SEKHON,
18 Q. SUN, M. TENAILLON, F. TIAN, J. WANG, X. XU, Z. ZHANG, S. KAEPLER,
19 J. ROSS-IBARRA, M. McMULLEN, E. BUCKLER, G. ZHANG, Y. XU, AND D. WARE.
20 2012. Maize hapmap2 identifies extant variation from a genome in flux. *Nat Genet* 44:
21 803–U238.
- 22 CLARK, R., T. WAGLER, P. QUIJADA, AND J. DOEBLEY. 2006. A distant upstream
23 enhancer at the maize domestication gene *tb1* has pleiotropic effects on plant and
24 inflorescent architecture. *Nat Genet* 38: 594–597.

- 1 COLOSIMO, P., K. HOSEMAN, S. BALABHADRA, G. VILLARREAL, M. DICKSON,
2 J. GRIMWOOD, J. SCHMUTZ, R. MYERS, D. SCHLUTER, AND D. KINGSLEY. 2005.
3 Widespread parallel evolution in sticklebacks by repeated fixation of ectodysplasin
4 alleles. *Science* 307: 1928–1933.
- 5 COOK, J., M. McMULLEN, J. HOLLAND, F. TIAN, P. BRADBURY, J. ROSS-IBARRA,
6 E. BUCKLER, AND S. FLINT-GARCIA. 2012. Genetic architecture of maize kernel
7 composition in the nested association mapping and inbred association panels. *Plant*
8 *Physiol* 158: 824–834.
- 9 COOP, G., D. WITONSKY, A. DI, RIENZO, AND J. PRITCHARD. 2010. Using
10 environmental correlations to identify loci underlying local adaptation. *Genetics* 185:
11 1411–1423.
- 12 DOEBLEY, J. 2004. The genetics of maize evolution. *Annu Rev Genet* 38: 37–59.
- 13 DOEBLEY, J. AND A. STEC. 1991. Genetic-analysis of the morphological differences
14 between maize and teosinte. *Genetics* 129: 285–295.
- 15 DOEBLEY, J. AND A. STEC. 1993. Inheritance of the morphological differences between
16 maize and teosinte: Comparison of results for two f_2 populations. *Genetics* 134:
17 559–570.
- 18 DOEBLEY, J., A. STEC, AND C. GUSTUS. 1995. *teosinte branched1* and the origin of
19 maize: Evidence for epistasis and the evolution of dominance. *Genetics* 141: 333–346.
- 20 DOEBLEY, J., A. STEC, AND L. HUBBARD. 1997. The evolution of apical dominance in
21 maize. *Nature* 386: 485–488.
- 22 DOYLE, J. AND J. DOYLE. 1990. A rapid total dna preparation procedure for small
23 quantities of fresh tissue. *Phytochemical Bulletin* 19: 11–15.

- 1 ELLSTRAND, N., L. GARNER, S. HEDGE, R. GUADAGNUOLO, AND L. BLANCAS. 2007.
2 Spontaneous hybridization between maize and teosinte. *Journal of Heredity* 98:
3 183–187.
- 4 ELLSTRAND, N., H. PRENTICE, AND J. HANCOCK. 1999. Gene flow and introgression
5 from domesticated plants into their wild relatives. *Annu Rev Ecol Syst* 30: 539–563.
- 6 FALUSH, D., M. STEPHENS, AND J. PRITCHARD. 2003. Inference of population structure
7 using multilocus genotype data: Linked loci and correlated allele frequencies. *Genetics*
8 164: 1567–1587.
- 9 FAO/IIASA/ISRIC/ISSCAS/JRC. 2012. Harmonized World Soil Database, version
10 1.2. FAO, Rome, Italy and IIASA, Laxenburg, Austria.
- 11 FEDER, J., S. BERLOCHER, J. ROETHELE, H. DAMBROSKI, J. SMITH, W. PERRY,
12 V. GAVRILOVIC, K. FILCHAK, J. RULL, AND M. ALUJA. 2003. Allopatric genetic
13 origins for sympatric host-plant shifts and race formation in rhagoletis. *P Natl Acad*
14 *Sci Usa* 100: 10314–10319.
- 15 FOX, J. AND S. WEISBERG. 2011. An R Companion to Applied Regression, vol. Second
16 Edition. Sage, Thousand Oaks, CA.
- 17 FUKUNAGA, N.-W. T. L. B. Z. Q. V. Y. F. M. B. K. L. L. D. J., K. ??? Genetic
18 diversity and population structure of teosinte. *Genetics* .
- 19 GALLAVOTTI, A., Q. ZHAO, J. KYOZUKA, R. MEELEY, M. RITTER, J. DOEBLEY,
20 M. PE, AND R. SCHMIDT. 2004. The role of barren stalk1 in the architecture of maize.
21 *Nature* 432: 630–635.
- 22 GERKE, J., J. EDWARDS, G. KE, J. ROSS-IBARRA, AND M. MCMULLEN. 2013. The
23 genomic impacts of drift and selection for hybrid performance in maize. *arXiv*
24 1307.7313.

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

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

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

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

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

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

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

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

- 1 KITANO, J., D. BOLNICK, D. BEAUCHAMP, M. MAZUR, S. MORI, T. NAKANO, AND
2 C. PEICHEL. 2008. Reverse evolution of armor plates in the threespine stickleback.
3 *Curr Biol* 18: 769–774.
- 4 KOVACH, M. AND S. MCCOUCH. 2008. Leveraging natural diversity: back through the
5 bottleneck. *Genome studies and Molecular Genetics* 11: 193–200.
- 6 LI, W. 2012. Tassels replace upper ears1 encodes a putative transcription factor that
7 regulates maize shoot architecture by multiple pathways. *PhD Dissertation* : 122.
- 8 LUKENS, L. AND J. DOEBLEY. 1999. Epistatic and environmental interactions for
9 quantitative trait loci involved in maize evolution. *Genet Res* 74: 291–302.
- 10 MALOOF, M., K. SOLIMAN, R. JORGENSEN, AND R. ALLARD. 1984. Ribosomal dna
11 spacer length polymorphisms in barley - mendelian inheritance, chromosomal location,
12 and population dynamics. *P Natl Acad Sci Usa* 81: 8014–8018.
- 13 MATSUOKA, Y., Y. VIGOUROUX, M. GOODMAN, G. SANCHEZ, E. BUCKLER, AND
14 J. DOEBLEY. 2002. A single domestication for maize shown by multilocus
15 microsatellite genotyping. *P Natl Acad Sci Usa* 99: 6080–6084.
- 16 OLSEN, K. AND B. GROSS. 2010. Genetic perspectives on crop domestication. *Trends in*
17 *Plant Science* 15: 529–537.
- 18 PIPERNO, D., A. RANERE, I. HOLST, J. IRIARTE, AND R. DICKAU. 2009. Starch grain
19 and phytolith evidence for early ninth millennium bp maize from the central balsas
20 river valley, mexico. *P Natl Acad Sci Usa* 106: 5019–5024.
- 21 PLANTINGA, T., S. ALONSO, N. IZAGIRRE, M. HERVELLA, R. FREGEL, J. VAN DER
22 MEER, M. NETEA, AND C. DE LA RUA. 2012. Low prevalence of lactase persistence in
23 neolithic south-west europe. *Eur J Hum Genet* 20: 778–782.

- 1 PYHÄJÄRVI, T., M. HUFFORD, AND J. ROSS-IBARRA. 2013. Complex patterns of local
2 adaptation in the wild relatives of maize. *Genome Biology and Evolution* 5: 1594–1609.
- 3 ROSS-IBARRA, J., P. MORRELL, AND B. GAUT. 2007. Plant domestication, a unique
4 opportunity to identify the genetic basis of adaptation. *P Natl Acad Sci Usa* 104:
5 8641–8648.
- 6 ROSS-IBARRA, J., M. TENAILLON, AND B. GAUT. 2009. Historical divergence and gene
7 flow in the genus *zea*. *Genetics* 181: 1399–1413.
- 8 ROZEN, S. AND H. SKALETSKY. 2000. Primer3 on the www for general users and for
9 biologist programmers. *Methods in Molecular Biology* : 365–386.
- 10 STUDER, A. AND J. DOEBLEY. 2012. Evidence for a natural allelic series at the maize
11 domestication locus *teosinte branched1*. *Genetics* 19: 951–958.
- 12 STUDER, A., Q. ZHAO, J. ROSS-IBARRA, AND J. DOEBLEY. 2011. Identification of a
13 functional transposon insertion in the maize domestication gene *tb1*. *Nat Genet* 43:
14 1160–U164.
- 15 THORNTON, K. 2003. libsequence: a c++ class library for evolutionary genetic analysis.
16 *Bioinformatics* 19: 2325–2327.
- 17 THURBER, C., M. REAGON, B. GROSS, K. OLSEN, Y. JIA, AND A. CAICEDO. 2010.
18 Molecular evolution of shattering loci in us weedy rice. *Mol Ecol* 19: 3271–3284.
- 19 TISHKOFF, S., F. REED, A. RANCIARO, B. VOIGHT, C. BABBITT, J. SILVERMAN,
20 K. POWELL, H. MORTENSEN, J. HIRBO, M. OSMAN, M. IBRAHIM, S. OMAR,
21 G. LEMA, T. NYAMBO, J. GHORI, S. BUMPSTEAD, J. PRITCHARD, G. WRAY, AND
22 P. DELOUKAS. 2007. Convergent adaptation of human lactase persistence in africa and
23 europe. *Nat Genet* 39: 31–40.

1 VAN HEERWAARDEN, J., J. DOEBLEY, W. BRIGGS, J. GLAUBITZ, M. GOODMAN,
2 J. GONZALEZ, AND J. ROSS-IBARRA. 2011. Genetic signals of origin, spread, and
3 introgression in a large sample of maize landraces. *P Natl Acad Sci Usa* 108:
4 1088–1092.

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

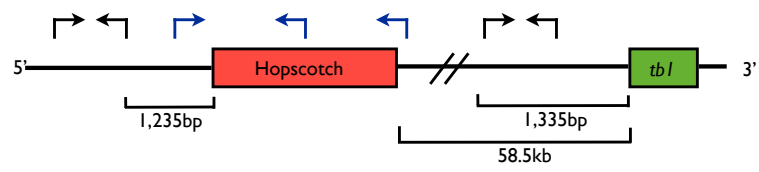
8 WHIPPLE, C., T. KEBROM, A. WEBER, F. YANG, D. HALL, R. MEELEY,
9 R. SCHMIDT, J. DOEBLEY, T. BRUTNELL, AND D. JACKSON. 2011. grassy tillers1
10 promotes apical dominance in maize and responds to shade signals in the grasses. *P*
11 *Natl Acad Sci Usa* 108: E506–E512.

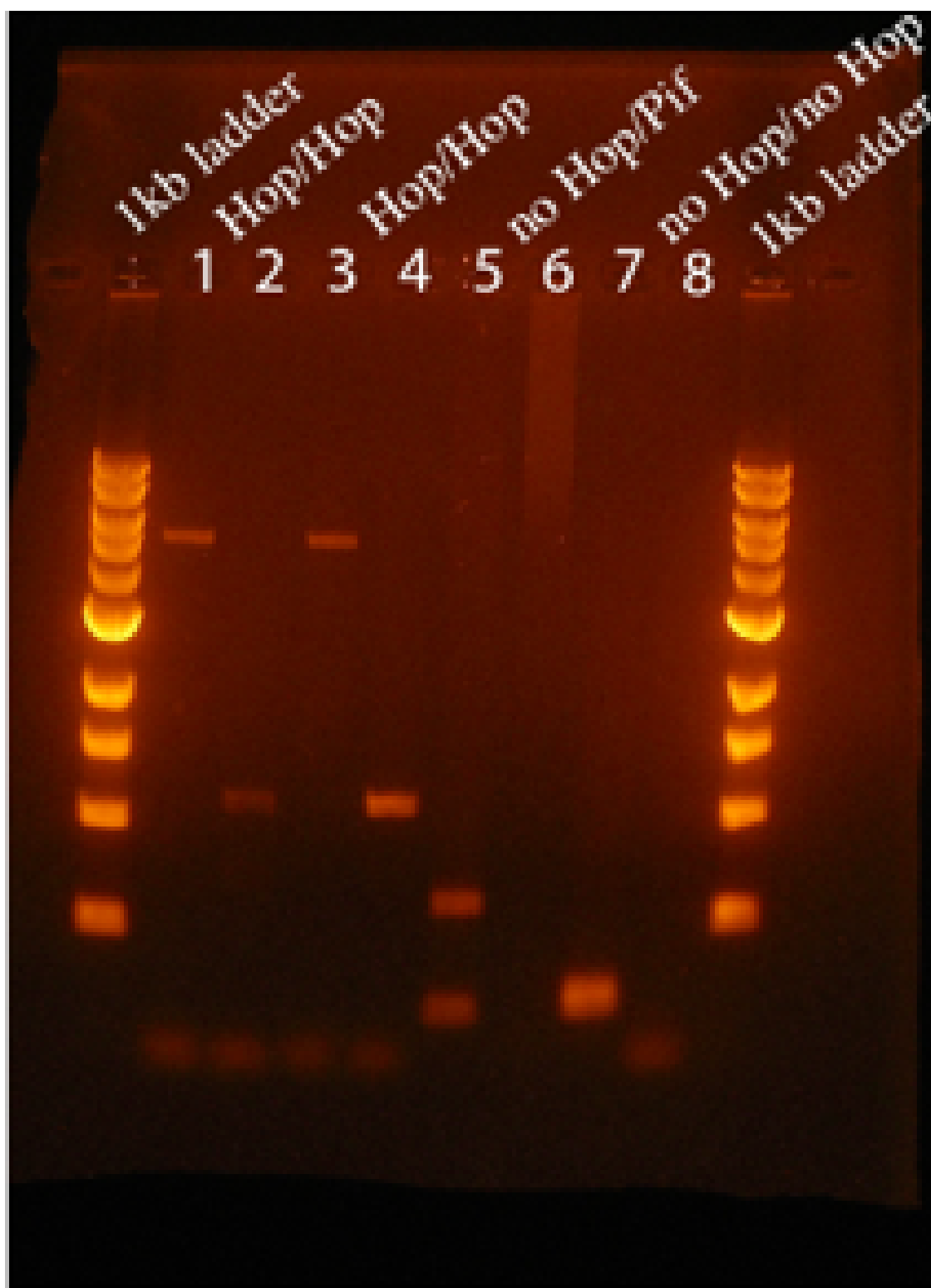
12 WHITTON J, A. D. S. A. R. L., WOLF DE. 1997. The persistence of cultivar alleles in
13 wild populations of sunflowers fiver generations after hybridization. *Theoretical and*
14 *Applied Genetcs* 95: 33–40.

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

17 ZHANG, L., Q. ZHU, Z. WU, J. ROSS-IBARRA, B. GAUT, S. GE, AND T. SANG. 2009.
18 Selection on grain shattering genes and rates of rice domestication. *New Phytol* 184:
19 708–720.

20 Supplementary Materials





| Accession | USDA Accession ID | Locality | Number alleles sampled | <i>Hopscotch</i> |
|-----------|-------------------|----------------------------|------------------------|------------------|
| 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 | 35 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 | <i>Hopscotch</i> Frequency |
|-----------|---------------------------|----------------------------|
| RIMMA0066 | 2 | 1 |
| RIMMA0075 | 2 | 1 |
| RIMMA0077 | 2 | 1 |
| RIMMA0079 | 2 | 1 |
| RIMMA0081 | 2 | 1 |
| RIMMA0084 | 2 | 1 |
| RIMMA0086 | 2 | 1 |
| RIMMA0088 | 2 | 1 |
| RIMMA0089 | 2 | 1 |
| RIMMA0090 | 2 | 1 |
| RIMMA0092 | 4 | 1 |
| RIMMA0094 | 4 | 1 |
| RIMMA0097 | 2 | 1 |
| RIMMA0099 | 2 | 1 |
| RIMMA0100 | 2 | 1 |
| RIMMA0101 | 2 | 1 |
| RIMMA0104 | 2 | 1 |
| RIMMA0108 | 2 | 1 |
| RIMMA0111 | 6 | 1 |
| RIMMA0115 | 2 | 1 |
| RIMMA0117 | 2 | 1 |
| RIMMA0130 | 2 | 1 |
| RIMMA0133 | 2 | 1 |
| RIMMA0134 | 2 | 1 |
| RIMMA0135 | 2 | 1 |
| RIMMA0142 | 2 | 0.5 |
| RIMMA0143 | 4 | 1 |
| RIMMA0146 | 4 | 1 |
| RIMMA0149 | 2 | 1 |
| RIMMA0152 | 2 | 1 |
| RIMMA0153 | 2 | 1 |
| RIMMA0154 | 2 | 1 |
| RIMMA0155 | 2 | 1 |

| PC1 | | PC2 | | PC3 | | PC4 | | PC5 | | PC6 | |
|----------|-------|---------|-------|---------|-------|---------|-------|---------|-------|---------|-------|
| Var | Rot | Var | Rot | Var | Rot | Var | Rot | Var | Rot | Var | Rot |
| bio1 | 0.146 | bio4 | 0.244 | prec7 | 0.287 | ts_clay | 0.41 | bio2 | 0.38 | bio10 | 0.086 |
| tmean11 | 0.146 | bio3 | 0.241 | prec8 | 0.276 | v_mod | 0.359 | sq4 | 0.328 | x_mod | 0.111 |
| tmean12 | 0.145 | bio7 | 0.241 | prec11 | 0.262 | ts_sand | 0.329 | ts_loam | 0.289 | sq3 | 0.207 |
| bio11 | 0.145 | prec6 | 0.237 | bio13 | 0.247 | bio15 | 0.272 | ts_sand | 0.266 | bio1 | 0.146 |
| tmax12 | 0.145 | sq7 | 0.218 | prec1 | 0.246 | prec4 | 0.259 | sq7 | 0.231 | v_mod | 0.359 |
| tmin5 | 0.145 | prec9 | 0.217 | bio16 | 0.242 | x_mod | 0.244 | bio18 | 0.213 | prec1 | 0.287 |
| tmean1 | 0.145 | sq3 | 0.207 | prec12 | 0.24 | prec3 | 0.226 | bio13 | 0.207 | bio1 | 0.146 |
| tmean2 | 0.145 | prec12 | 0.207 | bio19 | 0.238 | sq3 | 0.21 | prec11 | 0.183 | sq4 | 0.328 |
| tmin4 | 0.145 | bio12 | 0.204 | bio12 | 0.231 | prec5 | 0.21 | bio7 | 0.17 | sq3 | 0.207 |
| tmax1 | 0.145 | bio19 | 0.196 | prec2 | 0.222 | prec7 | 0.19 | bio16 | 0.163 | ts_sand | 0.329 |
| tmean4 | 0.145 | prec2 | 0.188 | bio18 | 0.221 | sq4 | 0.186 | bio4 | 0.157 | bio4 | 0.244 |
| tmin11 | 0.144 | prec1 | 0.185 | sq4 | 0.2 | bio3 | 0.185 | bio12 | 0.156 | prec1 | 0.287 |
| tmax11 | 0.144 | prec10 | 0.184 | prec9 | 0.18 | bio18 | 0.178 | bio3 | 0.155 | tmax1 | 0.145 |
| tmin12 | 0.144 | bio16 | 0.183 | prec10 | 0.171 | sq7 | 0.132 | prec6 | 0.154 | tmax1 | 0.145 |
| tmin2 | 0.144 | prec8 | 0.17 | prec5 | 0.161 | bio14 | 0.116 | x_mod | 0.152 | bio1 | 0.146 |
| tmean5 | 0.144 | prec5 | 0.165 | prec4 | 0.154 | bio13 | 0.099 | prec9 | 0.144 | tmax1 | 0.145 |
| tmean10 | 0.144 | bio14 | 0.158 | sq3 | 0.147 | bio16 | 0.095 | prec8 | 0.143 | bio1 | 0.146 |
| bio6 | 0.144 | bio13 | 0.151 | bio2 | 0.143 | prec8 | 0.09 | v_mod | 0.142 | ts_loam | 0.289 |
| tmax2 | 0.144 | bio17 | 0.149 | bio17 | 0.129 | bio7 | 0.077 | bio15 | 0.136 | ts_clay | 0.41 |
| tmean3 | 0.144 | prec3 | 0.144 | ts_loam | 0.127 | bio4 | 0.075 | prec7 | 0.112 | tmin1 | 0.143 |
| tmin1 | 0.143 | ts_clay | 0.141 | v_mod | 0.123 | bio2 | 0.074 | prec4 | 0.108 | tmin1 | 0.143 |
| tmin10 | 0.143 | bio2 | 0.129 | prec3 | 0.113 | prec2 | 0.074 | bio14 | 0.096 | prec1 | 0.287 |
| Altitude | 0.143 | prec7 | 0.108 | x_mod | 0.111 | bio19 | 0.068 | tmax7 | 0.093 | tmin1 | 0.143 |
| bio9 | 0.143 | tmax6 | 0.107 | bio14 | 0.099 | prec12 | 0.056 | tmax8 | 0.092 | tmin1 | 0.143 |
| tmin3 | 0.143 | x_mod | 0.106 | bio4 | 0.07 | ts_loam | 0.053 | prec1 | 0.091 | tmin1 | 0.143 |
| bio10 | 0.142 | bio15 | 0.098 | tmax3 | 0.067 | tmax12 | 0.047 | prec2 | 0.086 | tmean1 | 0.145 |
| tmax10 | 0.142 | ts_loam | 0.088 | ts_clay | 0.065 | bio17 | 0.047 | tmin11 | 0.086 | tmax1 | 0.145 |
| tmax3 | 0.142 | tmean6 | 0.085 | bio15 | 0.056 | bio9 | 0.043 | prec5 | 0.082 | tmax1 | 0.145 |
| tmax4 | 0.142 | tmin7 | 0.082 | tmax2 | 0.055 | tmax8 | 0.042 | bio17 | 0.082 | tmean1 | 0.145 |
| tmin6 | 0.142 | bio5 | 0.082 | tmean3 | 0.052 | tmax1 | 0.041 | tmin12 | 0.08 | bio1 | 0.146 |
| tmean9 | 0.141 | tmean7 | 0.081 | ts_sand | 0.05 | tmax5 | 0.039 | prec3 | 0.078 | tmin1 | 0.143 |
| tmin9 | 0.141 | prec4 | 0.08 | prec6 | 0.048 | tmax7 | 0.039 | tmax9 | 0.078 | prec1 | 0.287 |

| | | | | |
|-------------|---|---------|---------|---------|
| 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 |