Dear Dr.Vision and editorial staff at *PeerJ*,

We would like to thank you and two anonymous reviewers for your helpful comments and suggestions regarding our previously-submitted manuscript (#2014:12:3442:0:0:REVIEW). In the attached revision, we have addressed the majority of issues raised, and have included a detailed response to reviewers below. In our submitted manuscript, we report results of a genotyping assay surveying several hundred wild and domesticated maize individuals for presence of the domesticated maize allele at *tb1*, Sanger sequencing of a targeted subset of these samples to assess evidence of selection on the domesticated allele in wild populations, a phenotyping experiment to gauge the effects of the domesticated allele in a wild population, and reanalysis of existing genome-wide SNP data to place the signature of selection at this locus in a wider context. This manuscript reports a substantial amount of novel work and explores, in a meaningful way, the concept of domestication from standing variation and the potential ecological relevance of domesticated alleles in wild populations.

Sincerely,

Matthew Hufford and Jeffrey Ross-Ibarra

Response to Editor:

1) The reviewers comments seems to be largely in agreement and self-explanatory. Please be sure to address the comment regarding the validity of the PCR assays raised by Reviewer 2.
We have addressed comments from Reviewer 2 regarding the validity of the PCR assays. Sequencing of PCR products were not sequenced, but primer combinations for each of the 2 PCRs produced the expected sized bands. Primers were designed using previously sequenced teosinte and maize sequences provided by John Doebley, and were initially tested in both maize and teosinte for validity. We have clarified this point in the text text, line numbers here
and have added additional labeling to the gel image (Figure 2).

Response to Reviewer 1

Comments for the author

Populations of crop wild relatives often carry alleles associated with domestication traits at low to moderate frequencies. The presence of these alleles could potentially reflect post-domestication crop-to-wild gene flow or standing genetic variation that predates the domestication process. This study examines the frequency and phenotypic consequences of one of the best characterized crop domestication alleles, maize tb1, in populations of its wild ancestor, teosinte. The tb1 gene functions as a shade avoidance-mediated repressor of organ growth. In domesticated maize, a Hopscotch TE insertion in the tb1 promoter leads to enhanced gene expression and the loss of tillering and branching that characterizes domesticated maize. In this paper, the authors use PCR assays to survey for the presence of the TE insertion in a large sample of accessions, including maize, its wild ancestor teosinte, and highland teosinte (a separate subspecies less closely related to the domesticate). The study examines the potential ecological significance of the TE insertion allele through a Bayenv analysis, and population genetic analyses are used to confirm previous findings that the TE insertion predates maize domestication.

The paper is generally well written, and the conclusions are basically sound. Addressing the points below would improve the paper.

1) The terminology on accessions, individuals, and populations is confusing (e.g., 67-71). Please clarify the difference between individuals and accessions (e.g., line 70 is this referring to the number of individuals per accession? How are individuals of one accession related? Are these full sibs from the same maternal plant?..)

We have clarified this; accession refers to population; page 2, lines 62-66: "We sampled 1,110 individuals from 350 populations (247 maize landraces, 17 *mexicana* populations, and 86 *parviglumis* populations) and assessed the presence or absence of the *Hopscotch* insertion (Table S1 and Table S2, See Supplemental Materials with the online version of this article). Numbers of individuals sampled per population ranged from 1-43 for *parviglumis*, 1-35 for *mexicana*, and 1-18 for the maize landrace populations."

2) Is the Locality designation in Table S1 the same as population as used in the text? Please clarify and/or be consistent with this terminology.

Yes, we have changed the terminology in Table S1 to be consistent with the terminology used in the text.

3) Fig. 1 should indicate the boundaries of the Mexican states that are used in describing population locations in the text (Jalisco, Colima, etc.). The four focal populations from Jalisco state should also be labeled in this figure (i.e., San Lorenzo, La Mesa, Ejutla A, Ejutla B).

We feel that the addition of state lines makes the figure too busy and difficult to read. However, we have added labels to the 4 focal Jalisco populations (La Mesa (MSA), San Lorenzo (SLO), Ejutla A (EjuA), and Ejutla B (EjuB)).

I don't think state lines would make the figure too busy. Is the issue that you no longer have access to ArcGIS, LV? Jeff, do you think it's worth trying to redo this in R or do you think Todd will be fine with this response?

4) It would be helpful to include Fig. S1 in the main text of the paper and to label the two sequenced regions (Region 1 and 2) in the figure itself (not just in the caption).

Thank you for the suggestion. We have included Fig. S1 in the main manuscript, now labeled Fig. 1, and have added labels for the two sequenced regions, Region 1 and Region 2.

Hi guys, I'm having trouble getting the figures to renumber in order and haven't been able to figure out what's going on, if either of you knows how to fix or has a suggestion.

I fixed this

5) Fig. S2. Annotate this figure to show the size (bp) of each band in the 1 kb ladder. The caption should also define the abbreviations used in the labels.

We have edited this figure to be more clear.

6) What does No Hop/Pif indicate in Fig. S2? It doesn't have the three predicted bands of a TE heterozygote, and lane 6 seems to be empty altogether. Please clarify.

We have edited Fig. S2 (now Fig. S1) to be more clear and to include appropriate labels and explanations.

7) Line 118. Figure S1 should be cited here, not Table S1.

We are referring to Table S1 and Table S2. We have edited our reference to the tables to be more clear, page 8, lines 198-200: "The genotype at the *Hopscotch* insertion was confirmed with two PCRs for 837 individuals of the 1,100 screened (Table S1 and Table S2, See Supplemental Materials with the online version of this article)."

8) Line 150. Why were only 100 bootstrap replicates used for the NJ analysis? 1000 is more typical.

I do not know the answer to this. I don't think that 1000 bootstraps would give us really a difference.

if you still have the input file for this analysis, should be a quick fix and show we're doing everything we can to address reviewer concerns

agreed, should be easy to redo. while agree this is silly and i don't think 1000 will change anything, but since it's easy to do it's hard to argue against doing it.

I do not have the files as this is something that Tom did with the sequence data that he gathered

o.k., should we ask Tom for this? The only argument we can make for not doing this is that 100 bootstraps are commonly used, but this makes us sound pretty unaccommodating

9) Lines 164-175. It would be helpful to provide some more detail on the methods used in defining haplotypes in the STRUCTURE introgression analysis. If I understand correctly, SNPs within 5 kb windows were used to define haplotypes for that window. Were the haplotypes then used to define diploid genotypes for each individual in the STRUCTURE input file? If so, how was

phasing performed in defining the haplotypes? Is the output in Fig. 3 showing maize population assignment (i.e., membership coefficient values) for adjacent 5 kb windows? Was the STRUCTURE analysis performed only for the 8 Mb region shown in Fig. 3, or do the results presented in Table 4 reflect the entire chromosome?

Data were previously phased using the program fastPHASE (Pyhäjärvi et al. (2013); Scheet and Stephens (2006)). We have added clarifying statements to the text of this section, pages 5, lines 104-105. please go through and recheck these line numbers just prior to submitting revisions
Figure 4 is showing maize population assignment for adjacent 5 kb windows. The STRUCTURE analysis was performed for the entirety of chromosome 1; however, the figure only shows a zoomed in portion to highlight the section of chromosome 1 that contains the *tb1* region. The results in Table 4 reflect the entire chromosome.

11) Line 186. How does the 1-foot spacing of plants in the phenotyping experiment compare to conditions used in the experiments of Lukens and Doebley that documented density-dependent tillering? It could be worth bringing this up in the discussion of the negative phenotyping results (i.e., Fig 4 and Discussion, p. 13).

We have edited our discussion to include a brief mention of the work by Lukens and Doebley (1999), page 12, lines 344-346: "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. It is possible that the planting density of our seedlings (plants spaced 12 inches apart) was too high, leading to an overall decrease in tillering as previously seen in Lukens and Doebley (1999). This factor may have limited our capacity to observe variation in tillering index."

My guess is that later on our plants were crowded, but early on in the measurements they should not have been shading each other, and tillering then would not have been affected in the early measurements. What do you guys think?

I read the Lukens and Doebley paper and I think what you had previously was backwards...higher density = less tillering, not more. If we saw less overall tillering due to high density, this could have compromised our ability to detect significant differences between genotypes.

12) Line 223. Indicate here that these four populations are parviglumis and are from Jalisco state.

We have editied the text to provide further clarification, page 8, lines 211-213: "To investigate patterns of sequence diversity and linkage disequilibrium (LD) in the *tb1* region and any evidence of selection on this locus, we sequenced two small (<1kb) regions upstream of the *tb1* ORF in four populations from the Jalisco region."

13) Line 228. Why is Table S2 cited here? That table doesn't include any information on nucleotide diversity estimators.

We have removed reference to Table S2 from the citation.

14) Table 2 and the corresponding text (p. 9) should indicate which Tajima's D values, if any, are statistically significant. If none are statistically significant, then this should be pointed out in the text.

Okay. I am not sure that we have significance values for these.

We don't and it's kind of a strange/silly request...Jeff and I will deal with this

i can do the simple but dumb calculation if you guys want for response to reviewers, but disagree we should include in manuscript. we mention they are at the tail of the empirical distribution from wright et al don't we?

I agree this is a silly request and that we shouldn't include in the manuscript; yes, we indicate Taj.D for our pops is in the extreme tail based on comparison to both Wright et al. and Tiffin and Moeller

15) Lines 231-232. Suggested rewording: 'in all populations except La Mesa, where a slightly negative value suggests a slight excess of low frequency variants'

Thank you for the suggestion. We have reworded this sentence, page 9, lines 221-221: "in all populations except La Mesa, where a slightly negative value suggests a slight excess of low frequency

variants."

16) Lines 249-250. 'average r2 is slightly lower in the tb1 region..' Also, include these mean values as a row in Table 3.

The numbers reflected in Table 3 are averages.

17) Fig. 3 could be moved to online supplementary data, since the key information with respect to the tb1 region is already presented in Table 4.

We appreciate this suggestion but have decided to leave Figure 3 in the main text, as we feel it provides a nice visual representation for the numbers and take home message presented from the STRUCTURE analysis.

18) Discussion, lines 306-8. 'The Hopscotch allele is more prevalent in parviglumis than in mexicana in our sample, suggesting a different history of the allele amongst teosinte subspecies.' This inference is contradicted by the complete lack of differentiation between the subspecies for the TE (Fct = 0; Results, lines 213-217). This should be addressed. I don't see any evidence presented that the TE differences between the subspecies are statistically significant. Given the unbalanced geographical sampling of the two subspecies (see Table S1), I would also question whether these samples are appropriate for comparing TE frequencies in this way.

We have removed the comparisons made about the difference in frequencies of the *Hopscotch* allele between ssp. *parviglumis* and ssp. *mexicana*.

19) There are some sloppy inferences about selection in the Discussion and Conclusion 'e.g., lines 345-354, 388-389: there's no evidence that the Tajima's D values cited here are actually statistically significant. The statement and citations in lines 349-351 suggests that they likely are, but this should be backed up with tests of statistical significance. See comment 14 above.

Okay. What do you guys think of this? I don't think we have tests of significance for tajd

Jeff and I will address this

sloppy my ass. ¡censored¿ reviewer!

:)

Response to Reviewer 2

Basic reporting

1) In the Introduction, the authors give some background about teosinte, but do so mainly as the precursor of maize, not so much as a wild species. For instance, they do not give much information about the ecological differences between parviglumis and mexicana, or about what is known for this species in terms of population structure, population density or ecological behavior. This should be added. In particular, the authors state that the tb1 gene is suspected to play a role in shade avoidance, so information on population density should be highlighted: Is it similar in all teosinte populations? Does it vary with altitude? With other ecological data? Does it vary between parviglumis and mexicana?

We have added a few sentences discussing population density and distribution of both *Zea mays* ssp. *parviglumis* and *Zea mays* ssp. *mexicana* to the introduction. This addition is reflected in the text on page 2, lines 22-26: "Additionally, large, naturally-occurring populations of both *parviglumis* and *Zea mays* ssp. *mexicana* (hereafter, *mexicana*) can be found throughout Mexico (Wilkes, 1977; Hufford et al., 2013), with *parviglumis* distributed in the lowlands of Mexico and *mexicana* in the highlands. Furthermore, both *parviglumis* and *mexicana* can occur at high densities and genetic diversity of these taxa is estimated to be high (Hufford et al., 2012; Ross-Ibarra et al., 2009)."

2) The authors refer several times to a previous study they performed: Pyhjrvi et al., 2013. This is fine, but it is not always easy to understand which results are new and which come from this previous study. For instance, on line 158, it is stated 'we had whole genome SNP data come from Pyhjrvi et al., 2013. Similarly, Table 1 on line 217 should be moved right after 'populations (0.23) on line 216 to avoid confusion.

We have edited the reference to Table 1 and have edited the text to be more clear about which data were used where and how they were obtained: page 5, lines 97-100: "We used teosinte (ssp. *parviglumis* and ssp. *mexicana*) genotyping and covariance data from Pyhäjärvi et al. (2013) for BayEnv, with the *Hopscotch* insertion coded as an additional biallelic marker. SNP data from Pyhäjärvi et al. (2013) were obtained using the MaizeSNP50 BeadChip and Infinium HD Assay (Illumina, San Diego, CA, USA) and phased using the program fastPHASE (Scheet and Stephens, 2006). Environmental data were previously obtained for Pyhäjärvi et al. (2013) from `www.worldclim.org` and soil data were downloaded from the Harmonized World Soil Database (FAO/IIASA/ISRIC/ISSCAS/JRC, 2012) at `www.harvestchoice.org`." and page 9, lines 230-234: "We investigated the possibility of introgression as an explanation for the frequency of the *Hopscotch* allele in populations of teosinte using previously collected SNP data from (Pyhäjärvi et al., 2013)."

3) In the Results section, some titles like "Genotyping" "Sequencing" or "Phenotyping" are too vague, and sound more like Materials and Methods than to Results. They should be modified in order to highlight what question is addressed.

Thank you for the suggestion. We have edited our headings to be more detailed.

5) Paragraph "Neighbor joining trees" on lines 235-239 should be moved to the section "Evidence of introgression"

We have moved this to the "Evidence of Introgression" section.

Experimental design

6) The biological question is to characterize whether the Hopscotch transposable element at tb1 (which enhances expression of tb1) plays a role in the ecology of teosinte, especially in high-density populations. To do so, they characterize the distribution of this element in parviglumis, mexicana and maize landraces and they examine the phenotypic effects of the insertion in parviglumis. The authors state that they sampled 1,110 individuals from 350 accessions, with between 1-18 and 1-43 individuals per population. However, they do not provide explanation on (i) how these populations

were chosen and (ii) why a different number of individual was sampled per accession (visible also in suppl. Tables S1 and S2). This should be explained.

We sampled all individuals and populations that were available to us, which while it did not allow for even sample sizes across populations, did allow us to calculate *Hopscotch* frequency in a subset of populations, as well as the geographic distribution of the *Hopscotch* across many independent sampling locations. We have added a clarifying statement to the text regarding our sampling: page 4, lines 63-68: "We sampled all individuals and populations that were available to us, consisting of 1,110 individuals from 350 populations (247 maize landraces, 17 *mexicana* populations, and 86 *parviglumis* populations) and assessed the presence or absence of the *Hopscotch* insertion (Table S1 and Table S2, See Supplemental Materials with the online version of this article). Numbers of individuals sampled per population ranged from 1-43 for *parviglumis*, 1-35 for *mexicana*, and 1-18 for the maize landrace populations. Available samples did not allow us to sample evenly from populations, but did allow us to calculate *Hopscotch* frequency in a subset of populations, as well as elucidate the geographic distribution of the *Hopscotch* across multiple independent sampling sites."

5) In Suppl. Table S1, some of the parviglumis accessions are listed as "Breeders line". Can these be considered as natural populations? The authors should explain where they were originally collected. In Suppl. Table S2, the USDA (or other provider) ID corresponding to each accession should be provided, as it is done in suppl. Table S1.

I don't have access to this info. Can one of you guys help?

Jeff, can you help here?

6) The authors write that the PCR amplification to investigate presence of the Hopscotch element leads to two amplification products: one for the entire element, (5kb) and one for amplification of part of the element only (1.1kb). Why is this? How do the authors explain the origin of the 1.1kb band? The sentence with "and" on line 83, suggests that the two bands are produced in one single homozygous plant. If some accessions really amplify two bands, is the primer located in the LTR? If it is a typo and the "and" on line 83 should be replaced by a "or" is the element truncated in some accessions, leading to a 1.1kb band instead of the 5kb one? This should be clarified.

There are three primers that make up the two primer sets used to genotype individuals for the *Hopscotch* element. The primers are HopF, located upstream of the element; IntHopR, located within the LTR; and HopR, located downstream of the element. HopF and HopR are used as one reaction, resulting in either a 5kb product (the whole element) or a small band 300bp (sequence on either end of the element). HopF and IntHopR are used to amplify a portion of the element (1.1kb). Smaller bands preferentially amplify over larger bands when in the same reaction, so we chose to add the HopF/IntHopR PCR reaction as a way to get around PCR bias when amplifying both a larger (5kb) and smaller (300bp) band at the same time during the HopF/HopR reaction. A plant homozygous for the *Hopscotch* insertion will produce one 5kb band in the HopF/HopR reaction and one 1.1kb band in the HopF/HopIntR reaction. A heterozygous individuals will have two bands for the HopF/HopR reaction (one 5kb and one 300bp) as well as a 1.1kb band for the HopF/HopIntRR reaction, totaling 3 bands. The two reactions were run in separate wells on an agarose gel to allow for scoring.

We have added some clarifying text to the manuscript: page 4, lines 75-86: "Two PCRs were performed for each individual, one with primers flanking the *Hopscotch* (HopF/HopR) and one with a flanking primer (HopF) and an internal primer (HopIntR). Primer sequences are HopF, 5'-TCGTTGATGCTTTGATGGATGG-3'; HopR, 5'-AACAGTATGATTTCATGGGACCG-3'; and HopIntR, 5'-CCTCCACCTCTCATGAGATCC-3' (Fig. 1 and Fig. S1, See Supplemental Materials with the online version of this article). Homozygotes for the no-*Hopscotch* allele show a single band for absence of the element (∼300bp) produced by the HopF/HopR primer set, and 0 bands for the HopF/HopIntR primer set since they lack the LTR where the internal primer sequence is located. Homozygotes for the *Hopscotch* allele also show one band at 5kb for the HopF/HopR PCR product as well as one band at 1.1kb for the HopF/HopIntR PCR. Heterozygotes for the *Hopscotch* allele show three bands total; both a 300bp band and a 5kb band for the HopF/HopR PCR and a 1.1Kb band for the HopF/HopIntR PCR (Table S2, See Supplemental Materials with the online version of this article)."

7) From Figure S1, the middle primer seems to be right in the middle of the element, which suggests it was not designed in the LTR region. Is this true? In which part of the retrotransposon was the primer designed? The authors should add position of the retrotransposon LTRs and coding

regions on this figure.

The HopIntR primer is within the LTR element and was designed off of published sequence data provided by John Doebley.

8) In suppl. Figure S1, the names "HopF/HopR/HopIntR" should be added, as well as corresponding expected amplification sizes.

We have edited this figure to include primer names and added expected amplification sizes to the figure legend.

9) In suppl. Figure S2, band size of several bands (close in size to these amplified) should be indicated next to the ladder. Legend should explain better what the figure shows. It states "Genotypes are indicated at the top of the gel" What does this refer to? Numbers? "Hop/Hop" code? The figure should be completely relabeled, so that primer pair names and type of genotype (with presence of absence of Hopscotch) are clearly identified. What is "no Hop/Pif"? Why are there two bands on lane 5? Lane 6 has a weird smear. No primers are visible on lanes 5, 6 and 7.

We have edited the figure and the figure legend to be more clear.

10) Bottom gel resolution is poor. On the right for low molecular weight bands, it is difficult to assess that there is a single band (lane 7). A 1% gel and 1kb ladder are clearly not adequate for 300bp band detection.

We have edited the figure and the figure legend to be more clear. While the 300bp band is somewhat diffuse, we feel it is clear that there is a single band of the appropriate, expected size.

11) It is not stated whether PCR products were sequenced to check for correct amplification. Were some of the PCR products sequenced? If not, this should be done.

The only sequence data we have was when we sequenced a bit of the Pif to make sure that was what it was.

We did not sequence PCR products. We felt this was unnecessary based on the fact that primers were designed off of previous maize and teosinte sequence data. Furthermore, primers are unique to their location in the genome and do not BLAST with any confidence to any other location in the maize genome. PCRs produced bands of the expected sizes, which we felt was adequate.

12) The authors write: "Environmental data represent average values for the last several decades (climatic data) or are likely stable over time (soil data)". Does this mean that soil data was not averaged over the last decades? How was it estimated? The fact that soil parameters are "likely" stable over time actually quite depends on what we consider as "soil". Does this include micro-organisms? A bit more detail should be given.

Soil data are from the Harmonized World Soil Database `http://webarchive.iiasa.ac.at/Research/LUC/External-World-soil-database/HTML/` and include measurements for "organic Carbon, pH, water storage capacity, soil depth, cation exchange capacity of the soil and the clay fraction, total exchangeable nutrients, lime and gypsum contents, sodium exchange percentage, salinity, textural class and granulometry."

13) Figure 1: This figure should show 37 populations of parviglumis and 4 populations of mexicana (see page 8, lines 211 to 213). But the figure legends indicates only parviglumis. It should include mexicana populations, and they should be differentiated by an appropriate color code. In some cases, several circles seem to derive from one circle, suggesting they correspond to several populations from the same location. On the left hand side, it is unexpected to globally have a null frequency of "No Hopscotch" while several populations have a large fraction of "Hopscotch". This means that, within 25 miles, the frequency of the "Hopscotch" allele varies greatly. The authors should discuss this point. Names of the parviglumis populations used for the rest of the study (La Mesa, San Lorenzo, Ejutla1 and Ejutla2) should be indicated.

We only included *parviglumis* in this graph because we wanted to highlight the difference in frequency within the Jalisco region, even between populations that are geographically very close to each other. Reviewer 2 is correct in stating that there is a large difference in the frequency of the *Hopscotch* allele within 25 miles. Previous studies by Hufford (2010); Moeller et al. (2007) have

shown that populations in this area are highly differentiated and likely have different evolutionary histories.

We have edited the figure to include the names of the four populations in the Jalisco region that we intend to highlight: La Mesa, San Lorenzo, Ejulta A, and Ejutla B.

Validity of the findings

14) Results of the PCR amplifications are not well enough described and of high enough quality to be able to conclude whether the retrotransposon detection is valid or not (see "Experimental Design"). The authors should improve labelling of suppl. Figure S2 and, if needed, run the PCR amplicons on a more concentrated gel and with appropriate ladder.

We have edited the figure and the legend and corresponding text reference to be more clear about the expected banding patterns for the 3 genotypes.

15) Based on the association genetics study performed, the authors find an absence of association between the Hopscotch insertion and tillering index. At 40 days, they even find a weak but significant correlation, but in the unexpected direction (homozygotes for the Hopscotch insertion have a higher tillering). The authors discuss that this could be due to variation at other unlinked loci. However, they do not discuss on the effect of the environment on tillering. Considering the experiment was performed in greenhouse conditions while teosintes grow in much different environmental conditions, the genotype x environment interaction could differ greatly to this obtained in teosinte natural environmental conditions. This should be discussed.

The effects of *tb1* on tillering in maize have been verified a number of times in different environments Briggs et al. (2007); Clark et al. (2006); Studer and Doebley (2012); Studer et al. (2011). We feel that any effect of growing teosinte in a greenhouse setting as opposed to a a field trial would not be so large as to mask the effects of *tb1* on tillering.

References

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.

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.

FAO/IIASA/ISRIC/ISSCAS/JRC. 2012. Harmonized World Soil Database, version 1.2. FAO, Rome, Italy and IIASA, Laxenburg, Austria.

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

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

Hufford, M., P. Lubinsky, T. Pyhäjärvi, M. Devengenzo, N. Ellstrand, and J. Ross-Ibarra. 2013. The genomic signature of crop-wild introgression in maize. *PLoS Genetics* 9: e1003477.

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

Moeller, D. A., M. I. Tenaillon, and P. Tiffin. 2007. Population structure and its effects on patterns of nucleotide polymorphism in teosinte (zea mays ssp. parviglumis). *Genetics* 176: 1799–1809.

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

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

Scheet, P. and M. Stephens. 2006. A fast and flexible statistical model for large-scale population genotype data: applications to inferring missing genotypes and haplotypic phase. *American Journal of Human Genetics* 78: 629–644.

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

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

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