Recent demography drives changes in linked selection across the maize genome

Timothy M. Beissinger [*] [†] [‡], **Li Wang** [§] [¶], **Kate Crosby** [*], **Arun Durvasula** [*], **Matthew B. Hufford** [§], **and Jeffrey Ross-Ibarra** [*] [‖]

[*]Dept. of Plant Sciences, University of California, Davis, CA, USA,[†]US Department of Agriculture, Agricultural Research Service, Columbia, MO, USA,[‡]Division of Plant Sciences, University of Missouri, Columbia, MO, USA,[§]Department of Ecology, Evolution, and Organismal Biology, Iowa State University, Ames, IA, USA,[¶]Genome Informatics Facility, Iowa State University, Ames, IA, USA, and [‖]Genome Center and Center for Populat

Submitted to Proceedings of the National Academy of Sciences of the United States of Ame

The interaction between genetic drift and selection in shaping genetic diversity is not fully understood. In particular, a population's propensity to drift is typically summarized by its long-term effective population size (N_e) but rapidly changing population demographics may complicate this relationship. To better understand how changing demography impacts selection, we investigated linked selection in in the genomes of 23 domesticated maize and 13 wild maize (teosinte) individuals. We show that maize went through a domestication bottleneck with a population size of approximately 5% that of teosinte before it experienced rapid expansion post-domestication. We observe that hard sweeps on genic mutations are not the primary force driving maize evolution. As expected, a reduced population size during domestication decreased the efficiency of purifying selection to purge deleterious alleles from maize, but rapid expansion after domestication has increased the efficiency of purifying selection to levels exceeding those seen in teosinte. This final observation demonstrates that rapid demographic change can have wide-ranging impacts on diversity that conflict with would-be expectations based on long-term N_e.

The genetic diversity of populations is determined by a constant interplay between genetic drift and natural selection. Drift is a consequence of a finite population size and the random sampling of gametes every generation [1]. In contrast to the stochastic effects of drift, selection systematically alters allele frequencies by favoring particular alleles at the expense of others as a result of their effects on fitness. Researchers often seek to study drift by excluding potentially selected sites [2–4], or to study selection by focusing on site-specific patterns under the assumption that genome-wide diversity reflects primarily the action of drift [5].

Drift and selection do not operate independently to determine genetic variability, however, in large part because linkage allows the effects of selection to be wide-ranging [6–8]. Linked selection can take the form of hitch-hiking, when the frequency of a neutral allele changes as a result of positive selection at a physically linked site [6], or background selection, where diversity is reduced at loci linked to a site undergoing selection against deleterious alleles [9]. Recent work in a *Drosophila*, for example, has shown that virtually the entire genome is impacted by the combined effects of these processes [10–12].

The impact of linked selection, in turn, is heavily influenced by the effective population size (N_e), as the efficiency of natural selection is proportional to the product $N_e s$, where s is the strength of selection on a variant [8, 13–15]. The effective size of a population is not static, and nearly all species, including flies [16], humans [17], domesticates [18, 19], and non-model species [20] have experienced recent or ancient changes in N_e. Although much is known about how the long-term average N_e affects linked selection [13], relatively little is understood about the immediate effects of more recent changes in N_e on patterns of linked selection.

Because of its relatively simple demographic history and well-developed genomic resources, maize (*Zea mays*) represents an excellent organism to study these effects. Archae-

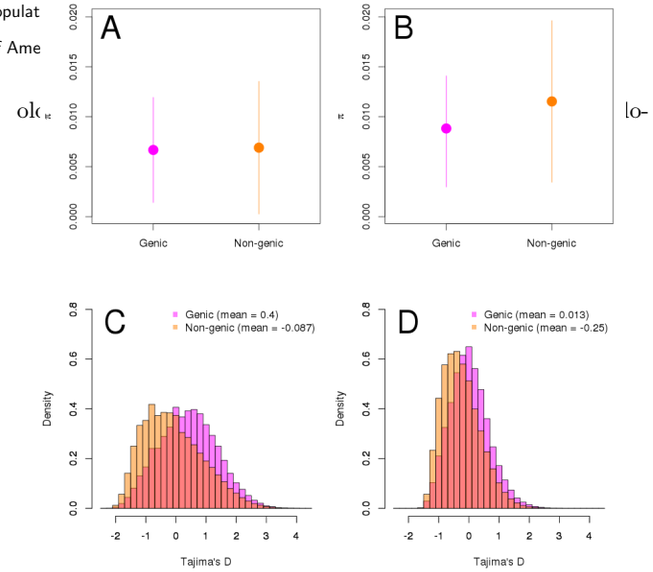

Fig. 1. **A** and **B** Show pairwise diversity π, while **C** and **D** depict and Tajimas D in 1kb windows from genic and nongenic regions of maize (A,C) and teosinte (B,D). Shown in A and B are means $\pm$ one standard deviation.

Significance

Both selection and demographic change play important roles in shaping diversity across the genome, but clear empirical examples of the interplay of these two forces are lacking. Here we document the combined effects of demography and linked selection on genome-wide diversity in domesticated maize and its wild ancestor teosinte. We estimate that maize underwent a bottleneck to $\approx 5\%$ of the size of the ancestral teosinte population, but that recent expansion has resulted in a maize population perhaps orders of magnitude larger than teosinte. We show that positive selection on new genic mutations has had relatively little effect on genetic diversity, but that selection against deleterious mutations has dramatically reduced diversity in and immediately around genes in both taxa. We find that the relative impact of selection depends on the age of the polymorphisms evaluated: while older polymorphisms in maize show more limited effects of linked selection, new mutations instead reflect its larger current size and more efficient selection. Our results demonstrate that a complete understanding of genome-wide patterns of diversity will require careful assessment of both demographic history and the effects of linked selection.

Reserved for Publication Footnotes

mestication began in Central Mexico at least 9,000 years bp [21, 22], and involved a population bottleneck followed by recent expansion [23–25]. Because of this simple but dynamic demographic history, domesticated maize and its wild ancestor teosinte can be used to understand the effects of changing N_e on linked selection. In this study, we leverage the maize-teosinte system to study these effects by first estimating the parameters of the maize domestication bottleneck using whole-genome resequencing data and then investigating the relative importance of different forms of linked selection on diversity in the recent and ancient past. We show that, while patterns of overall nucleotide diversity reflect long-term differences in Ne, recent growth following domestication qualitatively changes these effects, thereby illustrating the importance of a comprehensive understanding of demography when considering the effects of selection genome-wide.

Results

Patterns of diversity differ between genic and intergenic regions of the genome. To investigate how demography and linked selection have shaped patterns of diversity in maize and teosinte, we analyzed data from 23 maize and 13 teosinte genomes from the maize HapMap 2 and HapMap 3 projects [26, 27]. As a preliminary step, we evaluated levels of diversity inside and outside of genes across the genome. We find broad differences in genic and intergenic diversity consistent with earlier results [28](Figure 1). In maize, mean pairwise diversity (π) within genes was significantly lower than at sites at least 5 kb away from genes (0.00668 vs 0.00691, $p < 2\times10^{-44}$). Diversity differences in teosinte are even more pronounced (0.0088 vs. 0.0115, $p \approx 0$). Differences were also apparent in the site frequency spectrum, with mean Tajima's D positive in genic regions in both maize (0.4) and teosinte (0.013) but negative outside of genes (-0.087 in maize and -0.25 in teosinte, $p \approx 0$ for both comparisons). These observations suggest that diversity in genes is not evolving neutrally, but instead is reduced by the impacts of selection on linked sites.

Demography of maize domestication. After establishing that genic sites are not evolving neutrally, we used sites > 5kb from genes to estimate the parameters of a simple domestication bottleneck model (Figure 2). The most likely model estimates an ancestral population mutation rate of $\theta = 0.0147$ per bp, which translates to an effective population size of $N_a \approx 123,000$ teosinte individuals. We estimate that maize split from teosinte $\approx 15,000$ generations in the past, with an initial size of only $\approx 5\%$ of the ancestral N_a. After its split from teosinte, our model posits exponential population growth in maize, estimating a final modern effective population size of $N_m \approx 370,000$. Maize and teosinte have continued to exchange migrants after the population split, with gene flow between the populations estimated at $M_{tm} = 1.1 \times 10^{-5} \times N_a$ migrants per generation from teosinte to maize and $M_{mt} = 1.4 \times 10^{-5} \times N_a$ migrants from maize to teosinte.

In addition to our simple bottleneck model, we investigated two alternative approaches for demographic inference. First, we utilized genotyping data from more than 4,000 maize landraces [29] to estimate the modern maize effective population size using low frequency variants informative of population expansion. This analysis yields a much higher estimate of the modern maize effective population size at $N_m \approx 993,000$. Finally, we applied a model-free coalescent approach [30] using a subset of our samples. Though this analysis suggests non-equilibrium dynamics for teosinte not included in our initial model, it is nonetheless broadly consistent, identifying a clear domestication bottleneck followed by rapid population expansion in maize to an extremely large extant size of $\approx 10^9$ (Figure S2). Our assessment of the historical demography of maize and teosinte provides important context for subsequent analyses of linked selection.

Hard sweeps do not explain diversity differences. When selection increases the frequency of a new beneficial mutation, a signature of reduced diversity is left at surrounding linked sites [6]. To evaluate whether patterns of such "hard sweeps" could explain observed differences in diversity between genic and intergenic regions of the genome, we compared diversity around missense and synonymous substitutions between *Tripsacum* and either maize or teosinte. If a substantial proportion

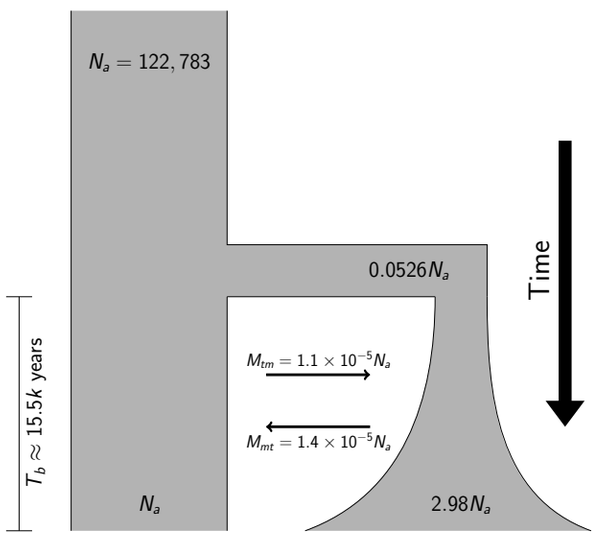

Fig. 2. Parameter estimates for a basic bottleneck model of maize domestication. See methods for details.

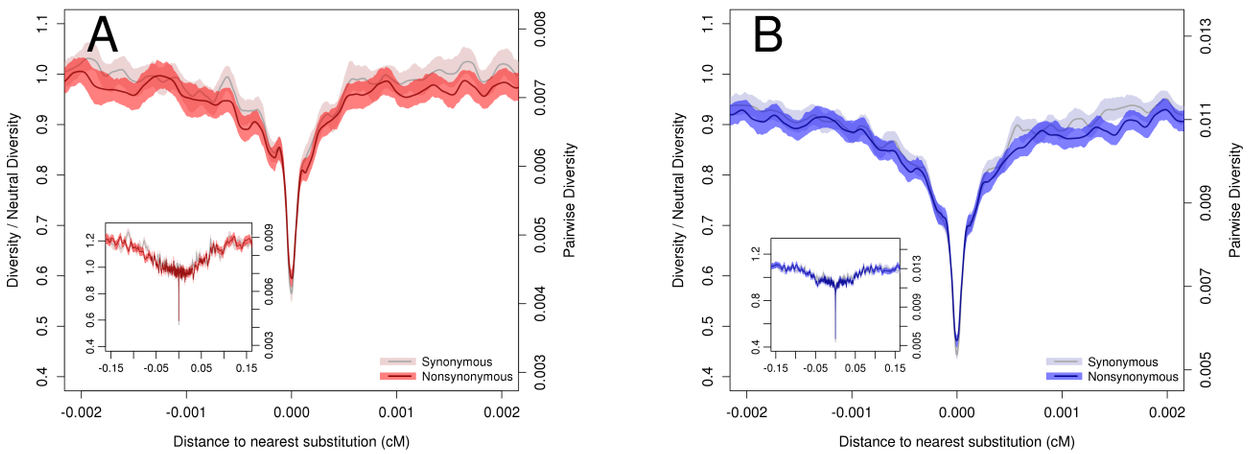

Fig. 3. Pairwise diversity surrounding synonymous and missense substitutions in **A** maize and **B** teosinte. Axes show absolute diversity values (right) and values relative to mean nucleotide diversity in windows $\geq 0.01cM$ from a substitution (left). Lines depict a loess curve (span of 0.01) and shading represents bootstrap-based 95% confidence intervals. Inset plots depict a larger range on the x-axis.

of missense mutations have been fixed due to hard sweeps, diversity around these substitutions should be lower than around synonymous substitutions. We observe this pattern around the causative amino acid substitution in the the maize domestication locus *tga1* (Figure S1), likely the result of a hard sweep during domestication [31, 32]. Genome-wide, however, we observe no differences in diversity between synonymous and missense substitutions in either maize or teosinte (Figure 3).

Previous analyses have suggested that this approach may have limited power because a relatively high proportion of missense substitutions will be found in genes that are under weak purifying selection and thus have higher genetic diversity [33]. To address this concern, we took advantage of genome-wide estimates of evolutionary constraint [34] calculated using genomic evolutionary rate profile (GERP) scores [35]. We then evaluated substitutions only in subsets of genes in the highest and lowest 10% quantile of mean GERP score, putatively representing genes under the strongest and weakest purifying selection. As expected, we see higher diversity around substitutions in genes under weak purifying selection, but we still find no difference between synonymous and missense substitutions in either subset of the data (Figure S3). Taken together, these data suggest hard sweeps do not play a major role in patterning genic diversity in either maize or teosinte.

Diversity is strongly influenced by purifying selection. In the case of purifying or background selection, diversity is reduced in functional regions of the genome via removal of deleterious mutations [9]. We investigated purifying selection in maize and teosinte by evaluating the reduction of diversity around genes. Pairwise diversity is strongly reduced within genes for both maize and teosinte (Figure 4A) but recovers quickly at sites outside of genes, consistent with the low levels of linkage disequilibrium generally observed in these subspecies [26, 36]. The reduction in relative diversity is more pronounced in teosinte, reaching lower levels in genes and occurring over a wider region.

Our previous comparison of synonymous and missense substitutions has low power to detect the effects of selection acting on multiple mutations or standing genetic variation, because in such cases diversity around the substitution may be reduced to a lesser degree [37,38]. Nonetheless, such "soft sweeps" are still expected to occur more frequently in functional regions of the genome and could provide an alternative explanation for the observed reduction of diversity at linked sites in genes. To test this possibility, we performed a genome-wide scan for selection using the H12 statistic, a method expected to be sensitive to both hard and soft sweeps [39]. Qualitative differences between maize and teosinte in patterns of diversity within and outside of genes remained unchanged even after removing genes in the top 20% quantile of H12 (Figure S6A). We interpret these combined results as suggesting that purifying selection has left a more pronounced signature in the teosinte genome due to the increased efficacy of selection resulting from differences in effective population size.

Population expansion leads to stronger purifying selection in modern maize. Motivated by the rapid post-domestication expansion of maize evident in our demographic analyses, we reasoned that low-frequency — and thus younger — polymorphisms might show patterns distinct from pairwise diversity. Singleton diversity around missense and synonymous substitutions (Figure S4) appears nearly identical to results from pairwise diversity (Figure 3), providing little support for a substantial recent increase in the number or strength of hard sweeps occurring in maize.

In contrast, we observe a significant shift in the effects of purifying selection: singleton polymorphisms are more strongly reduced in and near genes in maize than in teosinte, even after downsampling our maize data to account for differences in sample size (Figure 4B). This result is the opposite of the pattern observed for π, where teosinte demonstrated a stronger reduction of diversity in and around genes than did maize. As before, this relationship remained after we removed the 20% of genes with the highest H12 values (Figure S6). Finally, while direct comparison of pairwise and singleton diversity within taxa is consistent with non-equilibrium dynamics in teosinte, these too reveal much stronger differences in maize (Figure S5).

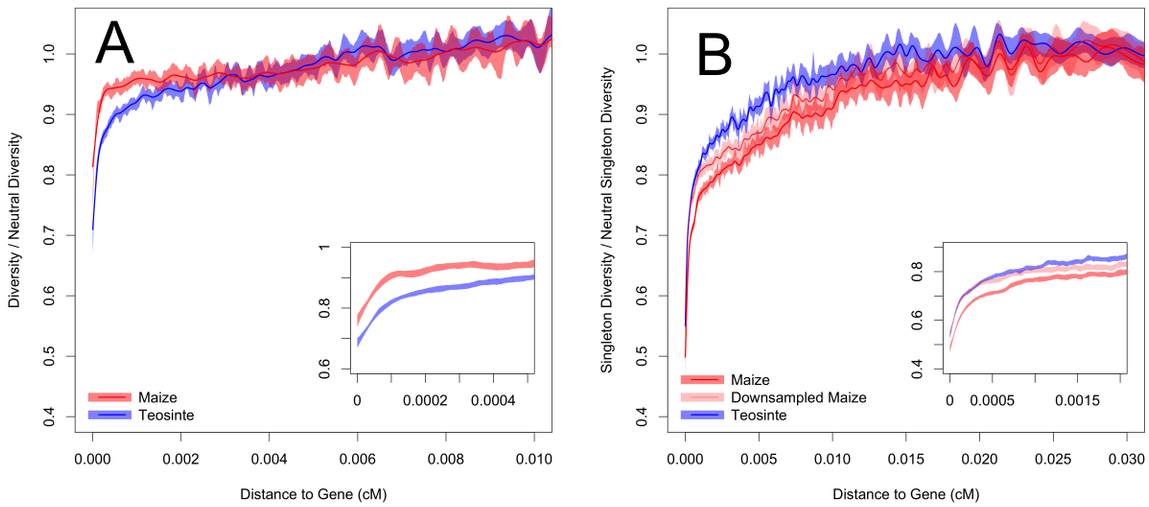

Fig. 4. Relative diversity versus distance to nearest gene in maize and teosinte. Shown are **A** pairwise nucleotide diversity and **B** singleton diversity. Relative diversity is calculated compared to the mean diversity in windows $\geq 0.01cM$ or $\geq 0.02cM$ from the nearest gene for pairwise diversity and singletons, respectively. Lines depict cubic smoothing splines with smoothing parameters chosen via generalized cross validation and shading depicts bootstrap-based 95% confidence intervals. Inset plots depict a smaller range on the x-axis.

Discussion

Demography of domestication. Although a number of authors have investigated the demography of maize domestication [23–25], these efforts relied on data only from genic regions of the genome and made a number of limiting assumptions about the demographic model. We show that diversity within genes has been strongly reduced by the effects of linked selection, such that even synonymous polymorphisms in genes are not representative of diversity at unconstrained sites. This implies that genic polymorphism data are unable to tell the complete or accurate demographic history of maize, but the rapid recovery of diversity outside of genes demonstrates that sites far from genes can be reasonably used for demographic inference. Furthermore, by utilizing the full joint SFS, we are able to estimate population growth, gene flow, and the strength of the domestication bottleneck without making assumptions about its duration.

One surprising result from our model is the estimated timing of domestication at $\approx 15,000$ years before present. While this appears to conflict with archaeological estimates [40], we emphasize that this estimate reflects the fact that the genetic split between populations likely preceded anatomical changes that can be identified in the archaeological record. We also note that our result may be inflated due to population structure, as our geographically diverse sample of teosinte may include populations diverged from those that gave rise to maize.

The estimated bottleneck of $\approx 5\%$ of the ancestral teosinte population seems low given that maize landraces exhibit $\approx 80\%$ of the diversity of teosinte [28], but our model suggests that the effects of the bottleneck on diversity are likely ameliorated by both gene flow and rapid population growth (Figure 2). Although we estimate that the modern effective size of maize is larger than teosinte, the small size of our sample reduces our power to identify the low frequency alleles most sensitive to rapid population growth [41], and our model is unable to incorporate growth faster than exponential. Both alternative approaches we employ estimate a much larger modern effective size of maize in the range of $\approx 10^6 - 10^9$, an order of magnitude or more than the current size of teosinte. Census data suggest these estimates are plausible: there are 47.9 million ha of open-pollinated maize in production [42], likely planted at a density of $\approx 25,000$ individuals per hectare [43]. Assuming the effective size is only $\approx 0.4\%$ of the census size (i.e. 1 ear for every 1000 male plants), this still implies a modern effective population size of more than four billion. While these genetic and census estimates are likely inaccurate, all of the evidence points to the fact that the effective size of modern maize is extremely large.

Hard sweeps do not shape genome-wide diversity in maize. Our findings demonstrate that classic hard selective sweeps have not contributed substantially to genome-wide patterns of diversity in maize, a result we show is robust to concerns about power due to the effects of background selection [33]. Although our approach ignores the potential for hard sweeps in noncoding regions of the genome, a growing body of evidence argues against hard sweeps as the prevalent mode of selection shaping maize variability. Among well-characterized domestication loci, only the gene *tga1* shows evidence of a hard sweep on a missense mutation [32], while several loci are consistent with "soft sweeps" from standing variation [44, 45] or multiple mutations [46]. Moreover, genome-wide studies of domestication [28], local adaptation [47] and modern breeding [48, 49] all support the importance of standing variation as primary sources of adaptive variation. Soft sweeps are expected to be common when $2N_e\mu_b \geq 1$, where μ_b is the mutation rate of beneficial alleles with selection coefficient s_b [38]. Assuming a mutation rate of 3×10^{-8} [50] and that on the order of $\approx 1-5\%$ of mutations are beneficial [51], this implies that soft sweeps should be common in both maize and teosinte for mutational targets $>> 10kb$ — a plausible size for quantitative traits or for regulatory evolution targeting genes with large up- or down-stream control regions [44, e.g.]. Indeed, many adaptive traits in both maize [52] and teosinte [53] are

highly quantitative, and adaptation in both maize [28] and teosinte [54] has involved selection on regulatory variation.

The absence of evidence for a genome-wide impact of hard sweeps in coding regions differs markedly from observations in *Drosophila* [55] and *Capsella* [56], but is consistent with data from humans [57, 58]. Comparisons of estimates of the percentage of nonsynonmyous substitutions fixed by natural selection [10, 56, 59, 60] give similar results. While larger long-term N_e likely explains some of this difference, we see little change in the importance of hard sweeps in genes in singleton diversity in modern maize (Figure S4, perhaps suggesting other factors may contribute to these differences. One possibility for example, is that if mutational target size scales with genome size, the larger genomes of human and maize may offer more opportunities for noncoding loci to contribute to adaptation, with hard sweeps on nonsynonymous variants then playing a relatively smaller role. Support from this idea comes from numerous cases of adaptive transposable element insertion modifying gene regulation in maize [44, 61–63] and studies of local adaptation that show enrichment for SNPs in regulatory regions in teosinte [54] and humans [64] but for nonsynonymous variants in the small *Arabidopsis* genome. Our results, for example, are not very different from those of [65], who find no differences in diversity around nonsynonymous and synonymous substitutions and estimate that as many as 80% of adaptive substitutions occur outside of genes. Future comparative analyses using a common statistical framework (e.g. [14]) and considering additional ecological and life history factors (c.f. [15]) should allow explicit testing of this idea.

Demography influences the efficiency of purifying selection. One of our more striking findings is that the impact of purifying selection on maize and teosinte qualitatively changed over time. We observe a more pronounced decrease in π around genes in teosinte than maize (Figure 4A), but the opposite trend when we evaluate diversity using singleton polymorphisms (Figure 4B). The efficiency of purifying selection is proportional to the effective population size [66], and these results are thus consistent with a domestication bottleneck and smaller long-term N_e in maize [23–25, 59] and our demographic analyses showing rapid expansion and a much larger modern N_e.

Although demographic change affects the efficiency of purifying selection, it may have limited implications for genetic load. Recent population bottlenecks and expansions have increased the relative abundance of rare and deleterious variants in domesticated plants [67, 68] and human populations out of Africa [41, 69], and such variants may play an important role in phenotypic variation [69–71]. Nonetheless, demographic history may have little impact on the overall genetic load of populations [72, 73], as decreases in N_e that allow weakly deleterious variants to escape selection also help purge strongly deleterious ones, and the increase of new deleterious mutations in expanding populations is mitigated by their lower initial frequency and the increasing efficiency of purifying selection [73–75].

Rapid changes in linked selection

Our results demonstrate that consideration of long-term differences in N_e cannot fully capture the dynamic relationship between demography and selection. While a number of authors have tested for selection using methods that explicitly incorporate or are robust to demographic change [60, 76, 77] and others have compared estimates of the efficiency of adaptive and purifying selection across species [78] or populations [79], previous analyses of the impact of linked selection on genome-wide diversity have relied on single estimates of the effective population size [14, 15]. Our results show that demographic change over short periods of time can quickly change the dynamics of linked selection: mutations arising in extant maize populations are much more strongly impacted by the effects of selection on linked sites than would be suggested by analyses using long-term effective population size. As many natural and domesticated populations have undergone considerable demographic change in their recent past, long-term comparisons of N_e are likely not informative about current processes affecting allele frequency trajectories.

Materials and Methods

BASH, R, and Python scripts. All scripts used for analysis are available in an online repository at https://github.com/timbeissinger/Maize-Teo-Scripts.

Plant materials. We made use of published sequences from inbred accessions of teosinte (Z. *mays* ssp. *parviglumis*) and maize landraces from the Maize HapMap3 panel as part of the Panzea project [26,27,80]. From these data, we removed 4 teosinte individuals that were not ssp. *parviglumis* or appeared as outliers in an initial principal component analysis conducted with the package adegenet [81] (Figure S7), leaving 13 teosinte and 23 maize that were used for all subsequent analyses (Table S8). We also utilized a single individual of (*Tripsacum dactyloides*) as an outgroup. All bam files are available at /iplant/home/shared/panzea/hapmap3/bam_internal/v3_bams_bwamem.

Physical and genetic maps. Sequences were mapped to the maize B73 version 3 reference genome [82] (ftp://ftp.ensemblgenomes.org/pub/plants/release-22/fasta/zea_mays/dna/) as described by [27]. All analyses made use of uniquely mapping reads with mapping quality score ≥ 30 and bases with base quality score ≥ 20; quality scores around indels were adjusted following [83]. We converted physical coordinates to genetic coordinates via linear interpolation of the previously published 1cM resolution NAM genetic map [84].

Estimating the site frequency spectrum. We estimated both the genome-wide site frequency spectrum (SFS) as well as a separate SFS for genic (within annotated transcript) and intergenic ($\geq 5kb$ from a transcript) regions. We used the biomaRt package [85, 86] of R [87] to parse annotations from genebuild version 5b of AGPv3. We estimated single population and joint SFS with the software ANGSD [88], including all positions with at least one aligned read in $\geq 80\%$ of samples in one or both populations. We assumed individuals were fully inbred and treated each line as a single haplotype. Because ANGSD cannot calculate a folded joint SFS, we first polarized SNPs using the maize reference genome and then folded spectra using $\delta\alpha\delta i$ [4].

Demographic inference. We used the software $\delta\alpha\delta i$ [4] to estimate parameters of a domestication bottleneck from the joint maize-teosinte SFS, using only sites $> 5kb$ from a gene to ameliorate the effects of linked selection. We modeled a teosinte population of constant effective size N_a, that at time T_b generations in the past gave rise to a maize population of size N_b which grew exponentially to size N_m in the present (Figure 2). The model includes migration of M_{mt} individuals each generation from maize to teosinte and M_{tm} individuals from teosinte to maize. We estimated N_a using $\delta\alpha\delta i$'s estimation of $\theta = 4N_a\mu$ from the data and a mutation rate of $\mu = 3 \times 10^{-8}$ [50]. We estimated all other parameters using 1,000 $\delta\alpha\delta i$ optimizations and allowing initial values between runs to be randomly perturbed by a factor of 2. Optimized parameters along with their initial values and upper and lower bounds can be found in table S9. We report parameter estimates from the optimization run with the highest log-likelihood.

We further made use of a large genotyping data set of more than 4,000 partially imputed maize landraces [29] to estimate the modern maize N_e from singleton counts. We filtered these data to include only SNPs with data in $\geq 1,500$ individuals, and then projected the SFS down to a sample of 500 individuals by sampling each marker without replacement 1,000 times according to the observed allele frequencies. We then estimated N_e from the data assuming $\mu = 3 \times 10^{-8}$ [50] and the relation $4N_e\mu = \frac{S}{L}$ [89], where where S is the total number of singleton SNPs and L is the total number of SNPs in the dataset.

As a final estimate of demography, we employed MSMC [30] to complement our model-based demographic inference. We used six each of maize and teosinte (BKN022, BKN025, BKN029, BKN030, BKN031, BKN033, TIL01, TIL03, TIL09, TIL10, TIL11 and TIL14), treating each inbred genome as a single haplotype. We

called SNPs in ANGSD [88] using a SNP p-value of $1e-6$ against a reference genome masked using SNPable (http://lh3lh3.users.sourceforge.net/snpable.shtml). We then removed heterozygous genotypes and filtered sites with a mapping quality < 30, a base quality < 20, or a $|log_2(\text{depth})| < 1$. We ran MSMC with pattern parameters $20 \times 2 + 20 \times 4 + 10 \times 2$.

Diversity. We made use of the software ANGSD [88] for diversity calculations and genotype calling. We calculated diversity statistics in maize and teosinte in 1 kb non-overlapping windows using filters as described above for the SFS. We used allele counts to estimate the number of singleton polymorphisms in each window, and used binomial sampling to create a second maize data set down-sampled to have the same number of samples as teosinte. We called genotypes in maize, teosinte, and *Tripsacum* at sites with a SNP p-value $< 10^{-6}$ and when the genotype posterior probability > 0.95. We identified substitutions in maize and teosinte as all sites with a fixed difference with *Tripsacum* and $\leq 20\%$ missing data. Substitutions were classified as synonymous, missense, or noncoding using the ensembl variant effects predictor [90]. For each window with $\geq 100bp$ of data we computed the genetic distance between the window center and the nearest synonymous and missense substitution as well as the genetic distance to the center of the nearest gene transcript.

Selection scan. We scanned the genome to identify sites that have experienced recent positive selection using the H12 statistic [39] in sliding windows of 200 SNPs with a step of 25 SNPs.

ACKNOWLEDGMENTS. We are indebted to Graham Coop and Simon Aeshbacher for their constructive input during this study. We thank Robert Bukowski and Qi Sun for providing early-access data from maize HapMap3. Funding was provided by NSF Plant Genome Research Project 1238014 and the USDA-Agricultural Research Service.

1. Dobzhansky, T & Pavlovsky, O. (1957) Evolution pp. 311–319.
2. Voight, B. F, Adams, A. M, Frisse, L. A, Qian, Y, Hudson, R. R, & Di Rienzo, A. (2005) Proceedings of the National Academy of Sciences of the United States of America 102, 18508–18513.
3. Luikart, G, England, P. R, Tallmon, D, Jordan, S, & Taberlet, P. (2003) Nature Reviews Genetics 4, 981–994.
4. Gutenkunst, R. N, Hernandez, R. D, Williamson, S. H, & Bustamante, C. D. (2009) PLoS genetics 5, e1000695.
5. Akey, J. M. (2009) Genome research 19, 711–722.
6. Smith, J. M & Haigh, J. (1974) Genetical research 23, 23–35.
7. Li, J, Li, H, Jakobsson, M, Li, S, SjÖDin, P, & Lascoux, M. (2012) Molecular Ecology 21, 28–44.
8. Slotte, T. (2014) Briefings in functional genomics 13, 268–275.
9. Charlesworth, B, Morgan, M, & Charlesworth, D. (1993) Genetics 134, 1289–1303.
10. Sella, G, Petrov, D. A, Przeworski, M, & Andolfatto, P. (2009) PLoS genetics 5, e1000495.
11. Elyashiv, E, Sattath, S, Hu, T. T, Strustovsky, A, McVicker, G, Andolfatto, P, Coop, G, & Sella, G. (2014) arXiv preprint arXiv:1408.5461.
12. Andolfatto, P. (2005) Nature 437, 1149–1152.
13. Cutter, A. D & Payseur, B. A. (2013) Nature Reviews Genetics 14, 262–274.
14. Corbett-Detig, R. B, Hartl, D. L, & Sackton, T. B. (2015) PLoS Biol 13, e1002112.
15. Leffler, E. M, Bullaughey, K, Matute, D. R, Meyer, W. K, Segurel, L, Venkat, A, Andolfatto, P, & Przeworski, M. (2012) PLoS Biol 10, e1001388.
16. Duchen, P, Živković, D, Hutter, S, Stephan, W, & Laurent, S. (2013) Genetics 193, 291–301.
17. Reich, D. E & Goldstein, D. B. (1998) Proceedings of the National Academy of Sciences 95, 8119–8123.
18. Hyten, D. L, Song, Q, Zhu, Y, Choi, I.-Y, Nelson, R. L, Costa, J. M, Specht, J. E, Shoemaker, R. C, & Cregan, P. B. (2006) Proceedings of the National Academy of Sciences 103, 16666–16671.
19. Consortium, B. H et al. (2009) Science 324, 528–532.
20. Ellegren, H. (2014) Trends in ecology & evolution 29, 51–63.
21. Smith, B. D. (1995) The emergence of agriculture. (Scientific American Library New York).
22. Matsuoka, Y, Vigouroux, Y, Goodman, M. M, Sanchez, J, Buckler, E, & Doebley, J. (2002) Proceedings of the National Academy of Sciences 99, 6080–6084.
23. Wright, S. I, Bi, I. V, Schroeder, S. G, Yamasaki, M, Doebley, J. F, McMullen, M. D, & Gaut, B. S. (2005) Science 308, 1310–1314.
24. Eyre-Walker, A, Gaut, R. L, Hilton, H, Feldman, D. L, & Gaut, B. S. (1998) Proceedings of the National Academy of Sciences 95, 4441–4446.
25. Tenaillon, M. I, U'Ren, J, Tenaillon, O, & Gaut, B. S. (2004) Molecular Biology and Evolution 21, 1214–1225.
26. Chia, J.-M, Song, C, Bradbury, P. J, Costich, D, de Leon, N, Doebley, J, Elshire, R. J, Gaut, B, Geller, L, Glaubitz, J. C, et al. (2012) Nature genetics 44, 803–807.
27. Bukowski, R, Guo, X, Lu, Y, Zou, C, He, B, Rong, Z, Wang, B, Xu, D, Yang, B, Xie, C, et al. (2015) bioRxiv p. 026963.
28. Hufford, M. B, Xu, X, Van Heerwaarden, J, Pyhäjärvi, T, Chia, J.-M, Cartwright, R. A, Elshire, R. J, Glaubitz, J. C, Guill, K. E, Kaeppler, S. M, et al. (2012) Nature genetics 44, 808–811.
29. Hearne, S, Chen, C, Buckler, E, & Mitchell, S. (2015) Unimputed gbs derived snps for maize landrace accessions represented in the seed-maize gwas panel (). Accessed: 2015-02-16.
30. Schiffels, S & Durbin, R. (2014) Nature genetics.
31. Wang, H, Nussbaum-Wagler, T, Li, B, Zhao, Q, Vigouroux, Y, Faller, M, Bomblies, K, Lukens, L, & Doebley, J. F. (2005) Nature 436, 714–719.
32. Wang, H, Studer, A. J, Zhao, Q, Meeley, R, & Doebley, J. F. (2015) Genetics pp. genetics–115.
33. Enard, D, Messer, P. W, & Petrov, D. A. (2014) Genome research 24, 885–895.
34. Rodgers-Melnick, E, Bradbury, P. J, Elshire, R. J, Glaubitz, J. C, Acharya, C. B, Mitchell, S. E, Li, C, Li, Y, & Buckler, E. S. (2015) Proceedings of the National Academy of Sciences 112, 3823–3828.
35. Davydov, E. V, Goode, D. L, Sirota, M, Cooper, G. M, Sidow, A, & Batzoglou, S. (2010) PLoS Comput Biol 6, e1001025.
36. Tenaillon, M. I, Sawkins, M. C, Anderson, L. K, Stack, S. M, Doebley, J, & Gaut, B. S. (2002) Genetics 162, 1401–1413.
37. Innan, H & Kim, Y. (2004) Proceedings of the National Academy of Sciences of the United States of America 101, 10667–10672.
38. Messer, P. W & Petrov, D. A. (2013) Trends in ecology & evolution 28, 659–669.
39. Garud, N. R, Messer, P. W, Buzbas, E. O, & Petrov, D. A. (2015) PLoS genetics 11, e1005004.
40. Piperno, D. R, Ranere, A. J, Holst, I, Iriarte, J, & Dickau, R. (2009) Proceedings of the National Academy of Sciences 106, 5019–5024.
41. Keinan, A & Clark, A. G. (2012) science 336, 740–743.
42. Program, T. M. (1999) Development, maintenance, and seed multiplication of open-pollinated maize varieties. (CIMMYT, Mexico, D.F.), 2 edition.
43. Baden, W. W & Beekman, C. S. (2001) American Antiquity pp. 505–515.
44. Studer, A, Zhao, Q, Ross-Ibarra, J, & Doebley, J. (2011) Nature genetics 43, 1160–1163.
45. Gallavotti, A, Zhao, Q, Kyozuka, J, Meeley, R. B, Ritter, M. K, Doebley, J. F, Pè, M. E, & Schmidt, R. J. (2004) Nature 432, 630–635.
46. Wills, D. M, Whipple, C. J, Takuno, S, Kursel, L. E, Shannon, L. M, Ross-Ibarra, J, & Doebley, J. F. (2013) PLoS Genet 9, e1003604.
47. Takuno, S, Ralph, P, Swarts, K, Elshire, R. J, Glaubitz, J. C, Buckler, E. S, Hufford, M. B, & Ross-Ibarra, J. (2015) Genetics.
48. van Heerwaarden, J, Hufford, M. B, & Ross-Ibarra, J. (2012) Proceedings of the National Academy of Sciences 109, 12420–12425.
49. Beissinger, T. M, Hirsch, C. N, Vaillancourt, B, Deshpande, S, Barry, K, Buell, C. R, Kaeppler, S. M, Gianola, D, & de Leon, N. (2014) Genetics 196, 829–840.
50. Clark, R. M, Tavaré, S, & Doebley, J. (2005) Molecular biology and evolution 22, 2304–2312.
51. Eyre-Walker, A & Keightley, P. D. (2007) Nature Reviews Genetics 8, 610–618.
52. Wallace, J, Larsson, S, & Buckler, E. (2014) Heredity 112, 30–38.
53. Weber, A. L, Briggs, W. H, Rucker, J, Baltazar, B. M, de Jesus Sánchez-Gonzalez, J, Feng, P, Buckler, E. S, & Doebley, J. (2008) Genetics 180, 1221–1232.
54. Pyhäjärvi, T, Hufford, M. B, Mezmouk, S, & Ross-Ibarra, J. (2013) Genome biology and evolution 5, 1594–1609.
55. Sattath, S, Elyashiv, E, Kolodny, O, Rinott, Y, & Sella, G. (2011) PLoS genetics 7, e1001302.
56. Williamson, R, Josephs, E, Platts, A, Hazzouri, K, Haudry, A, Blanchette, M, & Wright, S. (2014) PLoS genetics 10, e1004622–e1004622.
57. Hernandez, R. D, Kelley, J. L, Elyashiv, E, Melton, S. C, Auton, A, McVean, G, Sella, G, Przeworski, M, et al. (2011) science 331, 920–924.
58. Pritchard, J. K, Pickrell, J. K, & Coop, G. (2010) Current Biology 20, R208–R215.
59. Ross-Ibarra, J, Tenaillon, M, & Gaut, B. S. (2009) Genetics 181, 1399–1413.
60. Eyre-Walker, A & Keightley, P. D. (2009) Molecular biology and evolution 26, 2097–2108.
61. Castelletti, S, Tuberosa, R, Pindo, M, & Salvi, S. (2014) G3: Genes—Genomes— Genetics 4, 805–812.
62. Mao, H, Wang, H, Liu, S, Li, Z, Yang, X, Yan, J, Li, J, Tran, L.-S. P, & Qin, F. (2015) Nature Communications 6.
63. Yang, Q, Li, Z, Li, W, Ku, L, Wang, C, Ye, J, Li, K, Yang, N, Li, Y, Zhong, T, et al. (2013) Proceedings of the National Academy of Sciences 110, 16969–16974.
64. Fraser, H. B. (2013) Genome research 23, 1089–1096.
65. Halligan, D. L, Kousathanas, A, Ness, R. W, Harr, B, Eöry, L, Keane, T. M, Adams, D. J, & Keightley, P. D. (2013) PLoS Genetics 9, e1003995–14.
66. Kimura, M. (1984) The neutral theory of molecular evolution. (Cambridge University Press).

67. Günther, T & Schmid, K. J. (2010) Theoretical and Applied Genetics 121, 157–168.
68. Renaut, S & Rieseberg, L. H. (2015) Molecular biology and evolution p. msv106.
69. Coventry, A, Bull-Otterson, L. M, Liu, X, Clark, A. G, Maxwell, T. J, Crosby, J, Hixson, J. E, Rea, T. J, Muzny, D. M, Lewis, L. R, et al. (2010) Nature communications 1, 131.
70. Mezmouk, S & Ross-Ibarra, J. (2014) G3 (Bethesda, Md.) 4, 163–171.
71. Eyre-Walker, A. (2010) Proceedings of the National Academy of Sciences 107, 1752–1756.
72. Do, R, Balick, D, Li, H, Adzhubei, I, Sunyaev, S, & Reich, D. (2015) Nature genetics 47, 126–131.
73. Simons, Y. B, Turchin, M. C, Pritchard, J. K, & Sella, G. (2014) Nature genetics 46, 220–224.
74. Gazave, E, Chang, D, Clark, A. G, & Keinan, A. (2013) Genetics 195, 969–978.
75. Lohmueller, K. E. (2014) PLoS Genetics 10.
76. Chen, H, Patterson, N, & Reich, D. (2010) Genome research 20, 393–402.
77. Zeng, K & Charlesworth, B. (2010) Genetics 186, 1411–1424.
78. Popadin, K. Y, Nikolaev, S. I, Junier, T, Baranova, M, & Antonarakis, S. E. (2012) Molecular biology and evolution p. mss219.
79. Elyashiv, E, Bullaughey, K, Sattath, S, Rinott, Y, Przeworski, M, & Sella, G. (2010) Genome Research 20, 1558–1573.
80. Lemmon, Z. H, Bukowski, R, Sun, Q, & Doebley, J. F. (2014) PLoS Genet 10, e1004745.
81. Jombart, T & Ahmed, I. (2011) Bioinformatics 27, 3070–3071.
82. Schnable, P. S, Ware, D, Fulton, R. S, Stein, J. C, Wei, F, Pasternak, S, Liang, C, Zhang, J, Fulton, L, Graves, T. A, et al. (2009) science 326, 1112–1115.
83. Li, H. (2011) Bioinformatics 27, 2987–2993.
84. Glaubitz, J. C, Casstevens, T. M, Lu, F, Harriman, J, Elshire, R. J, Sun, Q, & Buckler, E. S. (2014) PLoS One 9, E90346.
85. Durinck, S, Spellman, P. T, Birney, E, & Huber, W. (2009) Nature protocols 4, 1184–1191.
86. Durinck, S, Moreau, Y, Kasprzyk, A, Davis, S, De Moor, B, Brazma, A, & Huber, W. (2005) Bioinformatics 21, 3439–3440.
87. R Core Team. (2014) R: A Language and Environment for Statistical Computing (R Foundation for Statistical Computing, Vienna, Austria).
88. Korneliussen, T. S, Albrechtsen, A, & Nielsen, R. (2014) BMC bioinformatics 15, 356.
89. Fu, Y.-X & Li, W.-H. (1993) Genetics 133, 693–709.
90. McLaren, W, Pritchard, B, Rios, D, Chen, Y, Flicek, P, & Cunningham, F. (2010) Bioinformatics 26, 2069–2070.

Supporting Information

first figure should be below header

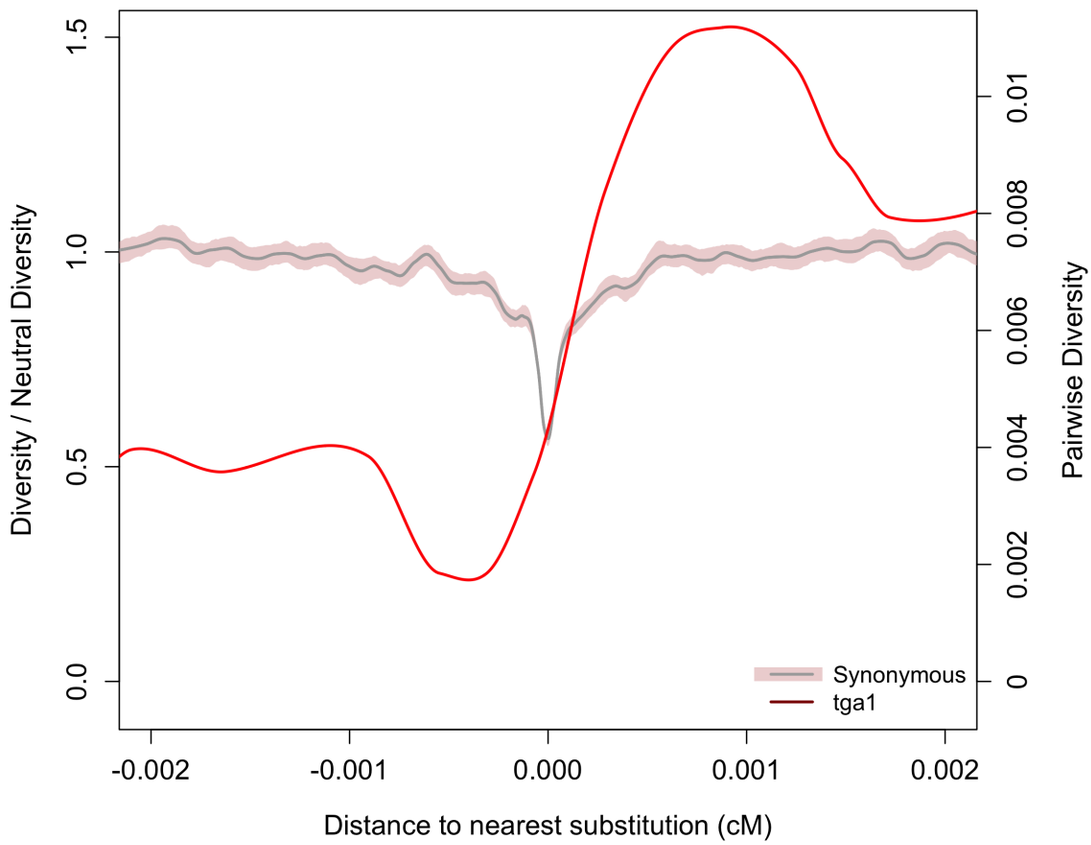

Fig. S1. Diversity surrounding the causitive substitution at the *tga1* locus.

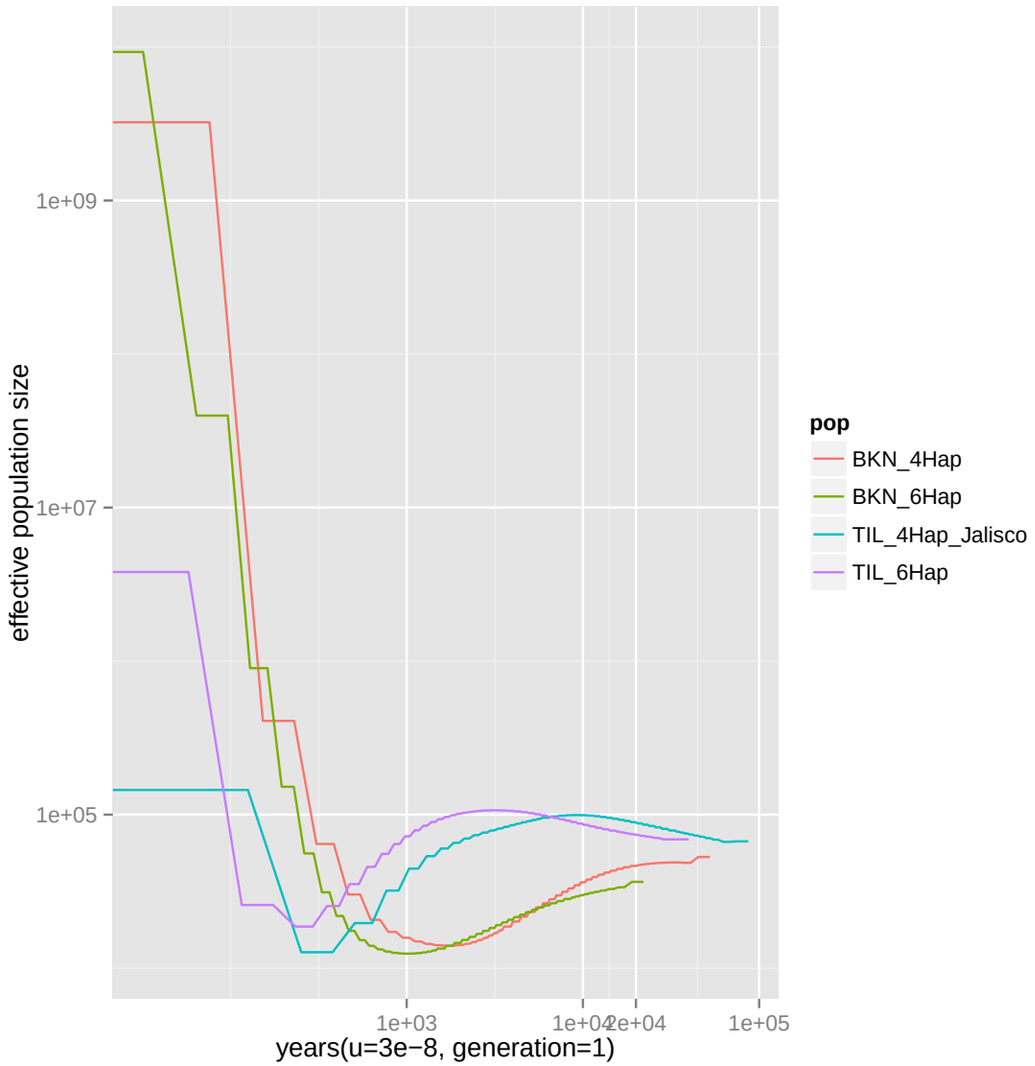

Fig. S2. need MSMC caption, remove gray background, fix axis labels and plot legend.

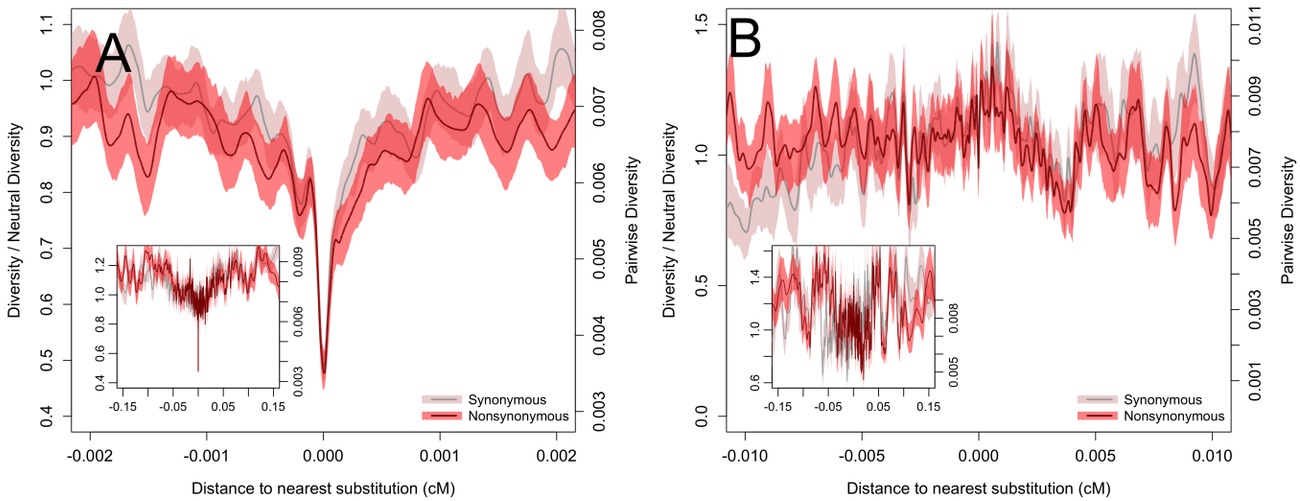

Fig. S3. Pairwise diversity surrounding synonymous and nonsynonymous substitutions in maize at **A** highly conserved or **B** unconserved sites. Bootstrap-based 95% confidence intervals are depicted via shading. Inset plots depict a larger range on the x-axis.

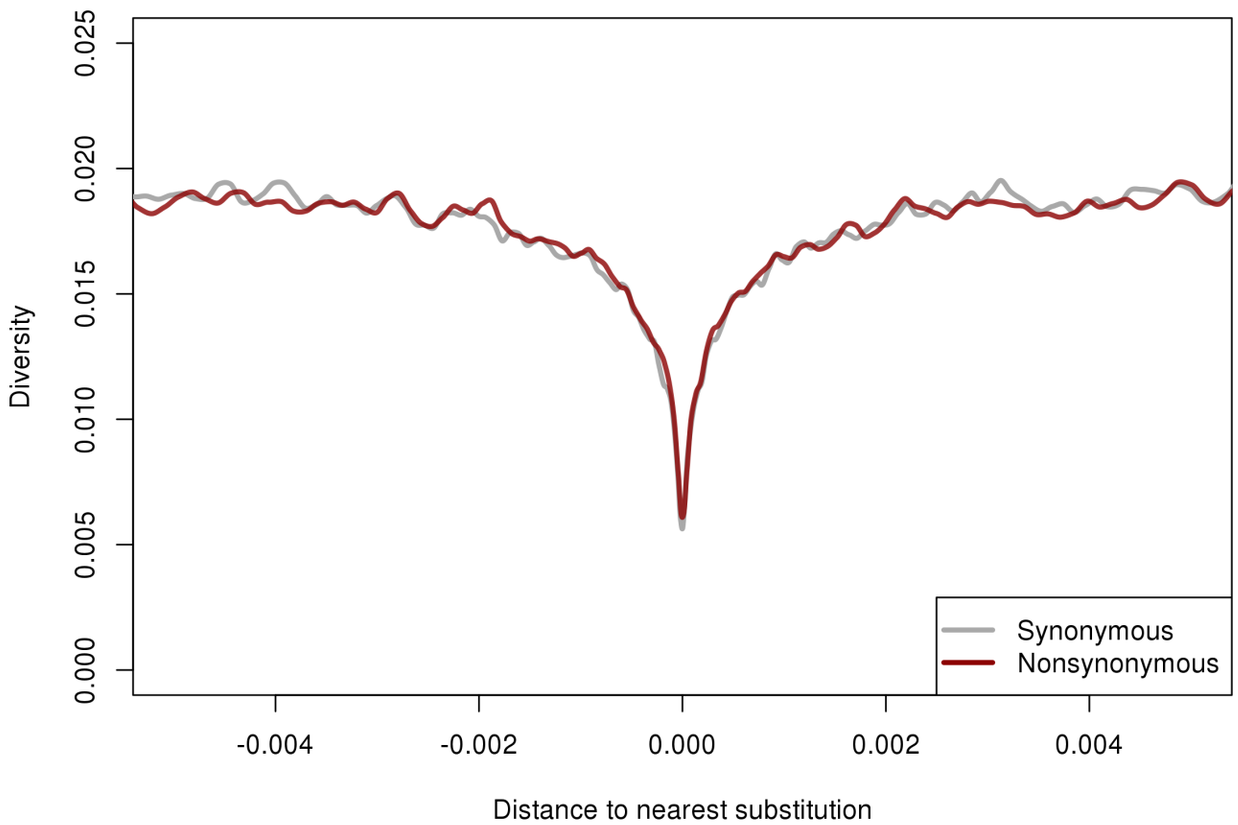

Fig. S4. Singleton diversity surrounding synonymous and nonsynonymous substitutions in maize.

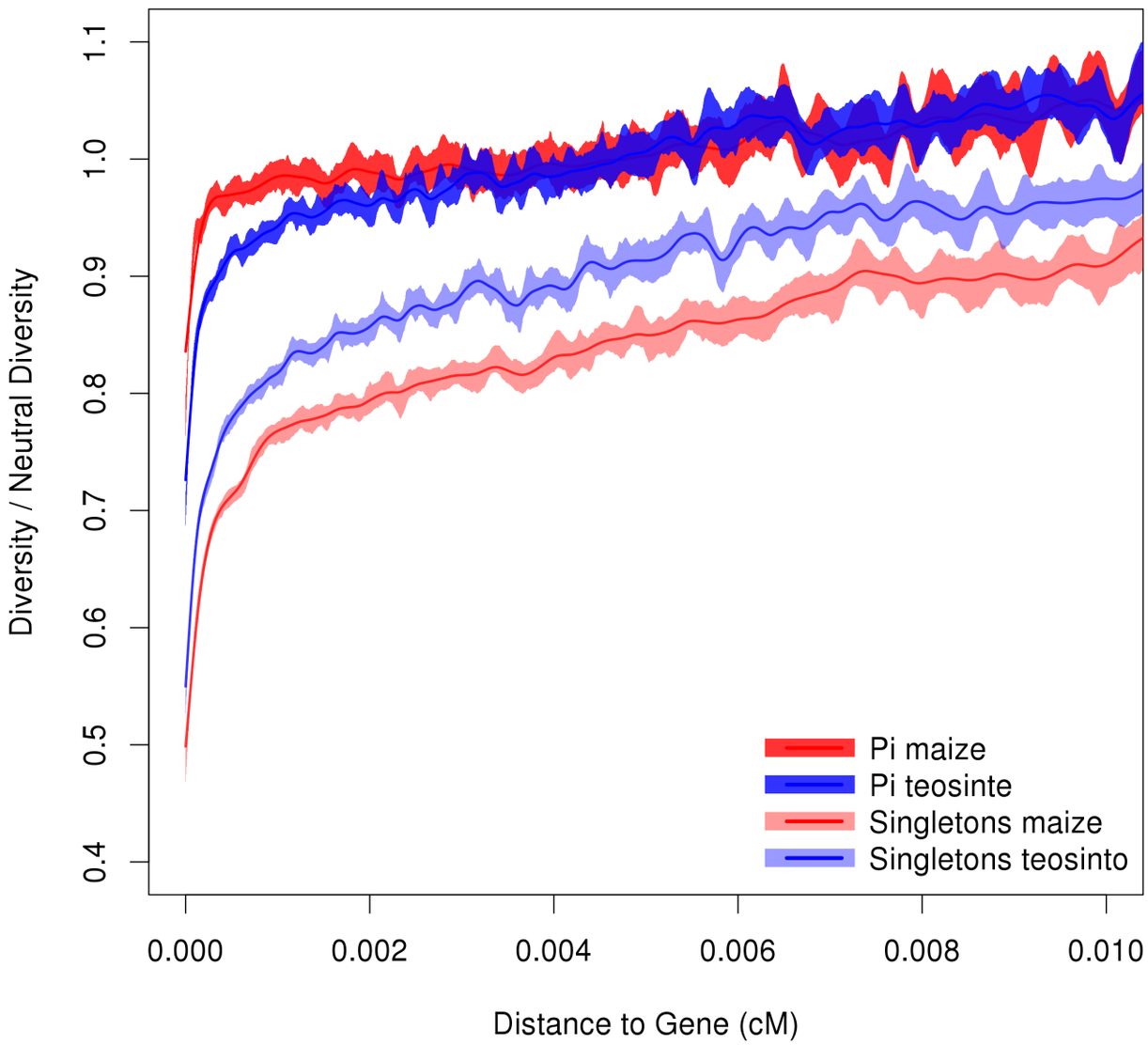

Fig. S5. Relative diversity versus distance to nearest gene in maize and teosinte. Relative diversity is calculated by comparing to the mean diversity in all windows $\geq 0.02cM$ from the nearest gene. Lines depict cubic smoothing splines with smoothing parameters chosen via generalized cross validation and shading depicts bootstrap-based 95% confidence intervals.

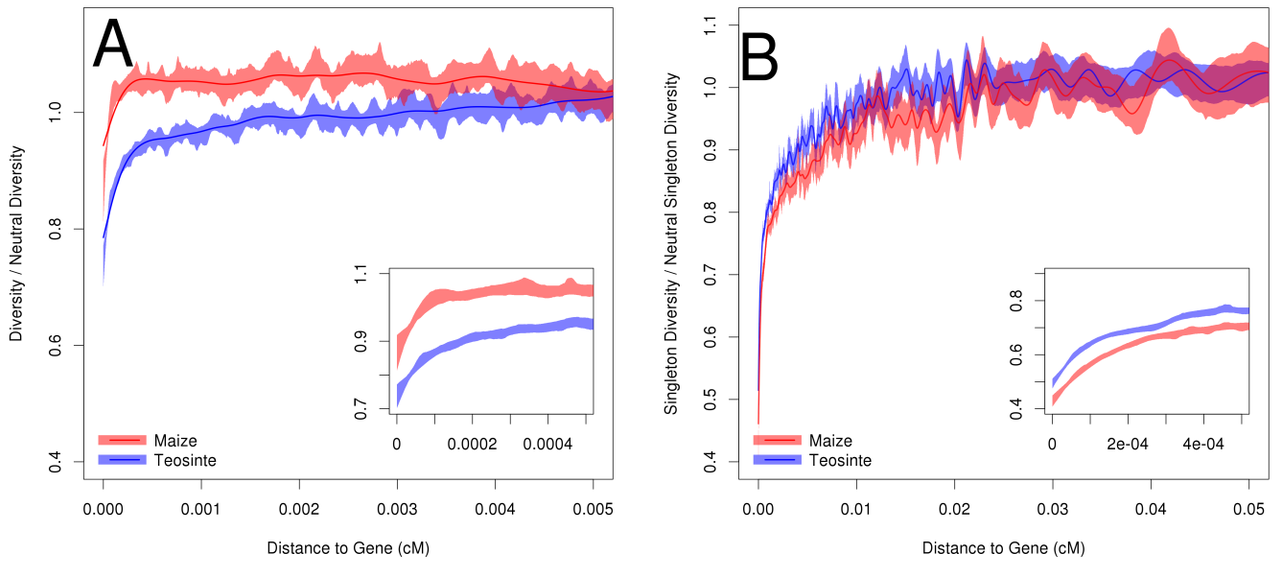

Fig. S6. Relative level of diversity versus distance to the nearest gene, in maize and teosinte, based on only sites that do not show evidence of hard or soft sweeps according to H12. Two measures of diversity were investigated. **A** displays pairwise diversity, which is most influenced by intermediate frequency alleles and therefore depicts more ancient evolutionary patterns, and **B** depicts singleton diversity, influenced by rare alleles and thus depicting evolutionary patterns in the recent past. Bootstrap-based 95% confidence intervals are depicted via shading. Inset plots depict a smaller range on the x-axis.

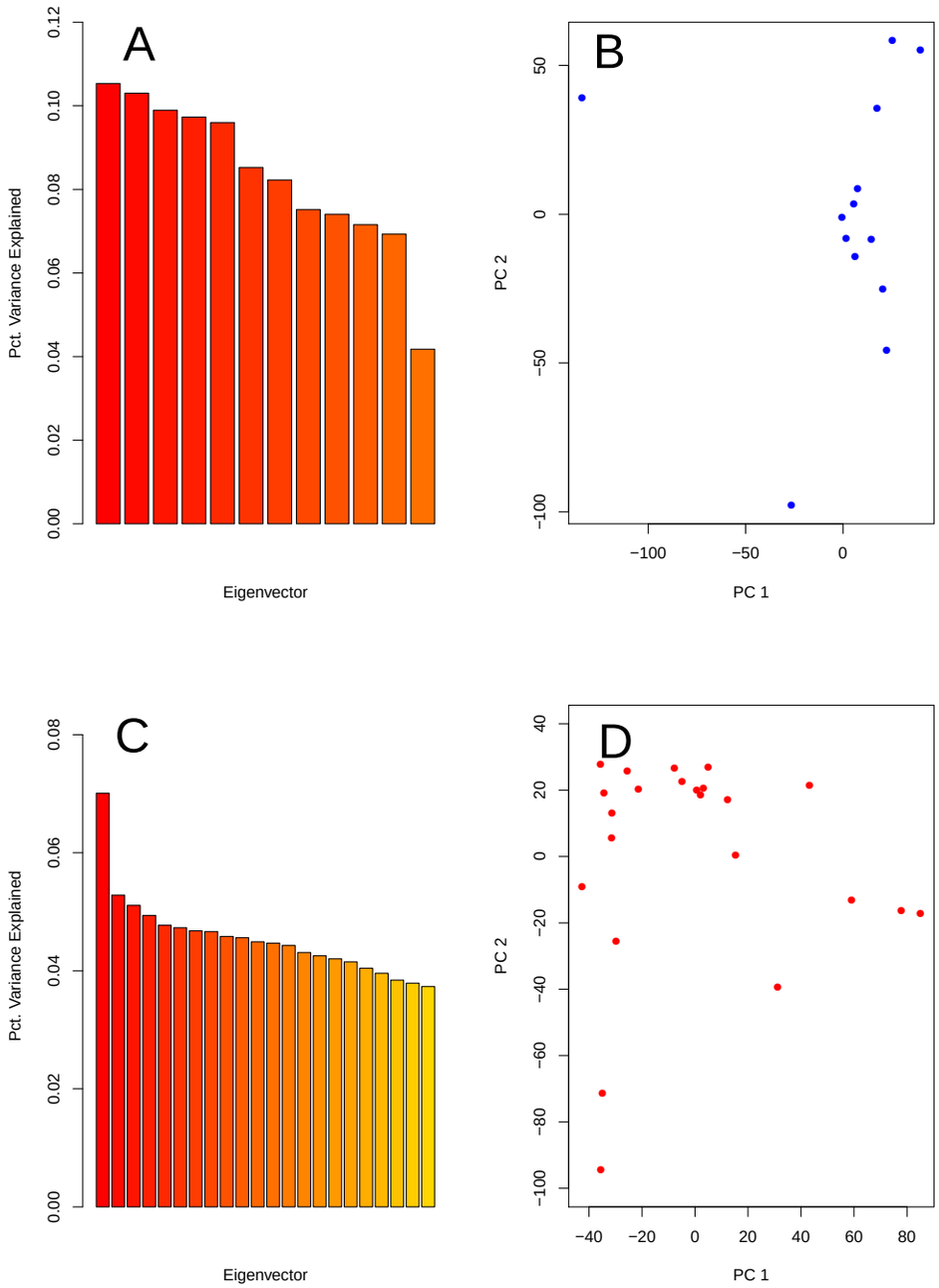

Fig. S7. Principal component analysis of teosinte and maize individuals to ensure that no close relatives were inadvertantly included in our study. Plots are based on a random sample of 10,000 SNPs. **A** displays the percentage of total variance explained by each principal component for teosinte, while **B** shows PC1 vs PC2 for all 13 teosinte individuals. Simlarly, **C** depicts the percentage of total variance explained by each principal component for maize, and **D** shows PC1 vs PC2 for all 23 maize individuals.

Maize	Teosinte
BKN009	TIL01
BKN010	TIL02
BKN011	TIL03
BKN014	TIL04-TIP454
BKN015	TIL07
BKN016	TIL09
BKN017	TIL10
BKN018	TIL11
BKN019	TIL12
BKN020	TIL14-TIP498
BKN022	TIL15
BKN023	TIL16
BKN025	TIL17
BKN026	
BKN027	
BKN029	
BKN030	
BKN031	
BKN032	
BKN033	
BKN034	
BKN035	
BKN040	

Fig. S8. A list of maize and teosinte individuals included in this study. Sequencing and details were previously described by [26]

Parameter	Initial value	Upper bound	Lower bound
$\frac{N_b}{N_a}$	0.02	1×10^{-7}	2
$\frac{N_m}{N_a}$	3	1×10^{-7}	200
$\frac{T_b}{2N_a}$	0.04	0	1
$\frac{M_{mt}}{N_a}$	1×10^{-10}	1×10^{-7}	0.001
$\frac{M_{tm}}{N_a}$	1×10^{-10}	1×10^{-7}	0.001

Fig. S9. Parameters, initial values, and boundaries used for model-fitting with $\delta\alpha\delta i$. Parameters are shown in the units utilized by $\delta\alpha\delta i$, although in the text simplified units are reported.

why are these shown as figures and not tables?