Not so local: the population genetics of convergent adaptation in maize and teosinte.

Silas Tittes[a,b,1], Anne Lorant[c], Sean McGinty[c], John F. Doebley[d], James B. Holland[e], Jose de Jesus Sánchez-González[f], Arun Seetharam[g], Maud Tenaillon[h] and Jeffrey Ross-Ibarra[a,b,i,1]

[a]Dept. of Evolution and Ecology, University of California, Davis, CA 95616, USA, [b]Center for Population Biology, University of California, Davis, CA 95616, USA, [c]Department of Plant Sciences, University of California, Davis, CA 95616, USA, [d]Laboratory of Genetics, University of Wisconsin – Madison, Madison, WI 53706 USA, [e]United States Department of Agriculture – Agriculture Research Service, Raleigh, NC 27695 USA and Department of Crop and Soil Sciences, North Carolina State University, [f]Centro Universitario de Ciencias Biológicas y Agropecuarias, Universidad de Guadalajara, Zapopan, Jalisco CP45110, México, [g]Department of Ecology, Evolution, and Organismal Biology; Genome Informatics Facility, Iowa State University, Ames, IA 50011, USA, [h]Génétique Quantitative et Evolution - Le Moulon, Université Paris-Saclay, INRAE, CNRS, AgroParisTech, 91190, Gif-sur-Yvette, France, [i]Genome Center, University of California, Davis, CA 95616, USA

ABSTRACT

What is the genetic architecture of local adaptation and what is the geographic scale that it operates over? We investigated patterns of local and convergent adaptation in five sympatric population pairs of traditionally cultivated maize and its wild relative teosinte (*Zea mays* subsp. *parviglumis*). We found that signatures of local adaptation based on the inference of adaptive fixations and selective sweeps are frequently exclusive to individual populations, more so in teosinte compared to maize. However, for both maize and teosinte, selective sweeps are frequently shared by several populations, and often between the subspecies. We were further able to infer that selective sweeps were shared among populations most often via migration, though sharing via standing variation was also common. Our analyses suggest that teosinte has been a continued source of beneficial alleles for maize, post domestication, and that maize populations have facilitated adaptation in teosinte by moving beneficial alleles across the landscape. Taken together, out results suggest local adaptation in maize and teosinte has an intermediate geographic scale, one that is larger than individual populations, but smaller than the species range.

KEYWORDS Local Adaptation, Convergent Adaptation, Maize, Teosinte

Introduction

As populations diverge and occupy new regions, they become locally adapted to the novel ecological conditions that they encounter. Decades of empirical work have carefully documented evidence for local adaptation, including the use of common garden and reciprocal transplant studies demonstrating that populations express higher fitness in their home environment (Clausen *et al.* 1948) as well as quantitative genetic approaches that show selection has acted on individual traits to make organisms better suited to their ecological conditions (Savolainen *et al.* 2013). It is clear from these studies that local adaptation is pervasive in natural populations.

One important but understudied aspect of local adaptation is its geographic scale. Empirical studies have documented adaptation at multiple scales, from microgeographic differentiation among mesic and xeric habitats along a single hillside (Hamrick and Allard 1972) to regional (Lowry *et al.* 2008; Whitehead *et al.* 2011) and even global scales (Colosimo *et al.* 2005). A key factor determining the geographic scale of local adaptation is the distribution of the biotic and abiotic challenges to which organisms are adapting, as these features place limits on the locations over which an allele remains beneficial. Environmental features overlap with each other to varying degrees (Tuanmu and Jetz 2015). The degree of overlap between environmental features may be important if mutations are pleiotropic, as an allele may not be beneficial when integrating its effect over multiple selective pressures (Chevin *et al.* 2010).

The geographic scale of local adaptation depends, too, on population structure. Gossmann *et al.* (2010) showed that the estimates of the proportion of new mutations fixed by natural selection across a number of plant species tended to overlap with zero, suggesting there is little evidence for adaptation at non-synonymous sites. One potential explanation raised by the authors for this surprising finding was that natural populations are often structured, such that very few adaptations would be expected to be common over the entire range of the species' distribution. Indeed, even when selective pressures are shared across populations, structure can hinder a species' adaptation by limiting the spread of beneficial alleles across the species range (Bourne *et al.* 2014). Consistent with this, Fournier-Level *et al.* (2011) conducted a continent-scale survey across strongly structured populations of *Arabidopsis thaliana*, finding that alleles which increase fitness tended to occur over a restricted geographic scale. Despite the progress made, it remains unclear if the scales identified in *Arabidopsis* are common to local adaptation across the entire genome, and how similar the general

[1] For correspondence, Dept. of Evolution and Ecology, University of California, Davis, CA, USA E-mail: stittes@ucdavis.edu; rossibarra@ucdavis.edu

patterns are across taxa.

The majority of local adaptation studies are motivated by conspicuous differences in the phenotypes or environments of two or more populations. As such, many instances of local adaptation that are occurring, as well as the underlying beneficial mutations being selected, may go overlooked. This hinders our ability to draw more general conclusions about the overall frequency and impact of local adaptation on a given population's evolutionary history. Efforts to more systematically survey evidence of local adaptation have led to "reverse ecology" approaches, where signatures of adaptive evolution are first identified from genomic sequencing data, and are then related to the traits, history, or environmental conditions of the populations in which they occur (Li *et al.* 2008; Levy and Borenstein 2012). Given their ability to survey numerous loci and populations simultaneously, reverse ecology and other population genetic approaches are well-suited to investigating the geographic scale of local adaptation. While the use of genome-scale reverse ecology approaches is increasing (Hoban *et al.* 2016; Bragg *et al.* 2015), sampling across multiple populations and explicit evaluation of the scale of local adaptation are still uncommon.

Using reverse ecology, we can compare the observed distribution of beneficial alleles across multiple populations. The patterns of selective sweeps and adaptive fixations that are exclusive to or shared among multiple populations can be used to measure a beneficial allele's geographic extent, which is influenced by the factors outlined above. If we infer multiple structured populations have fixed the same beneficial allele, this suggests that pleiotropy has not disrupted the adaptive value of the allele across environments or that the populations share a sufficiently similar set of selective pressures. Assessment of the relative frequency and geographic extent of unique and shared beneficial alleles thus allows us to quantify the scale of local adaptation. Additionally, when multiple populations do share an adaptive allele, we can infer the mode by which sharing occurred (Lee and Coop 2017), providing further insights about the environmental and genetic context of each adaptation as well as the processes underlying allele sharing among populations.

Motivated to improve our understanding about the genetic basis of local adaptation and its geographic scale, we set out to use the reverse ecology approach to understand patterns of adaptation via selective sweeps in multiple discrete populations of domesticated maize *Zea mays* ssp. *mays* and its wild relative teosinte *Zea mays* ssp. *parviglumis* growing across their native range in Mexico. *Zea mays* is an annual grass, native to southern Mexico. Maize was domesticated $\approx 9,000$ years ago (Piperno *et al.* 2009) from its common ancestor with the extant annual grass teosinte, but traditional open-pollinated populations maintain an extremely large population size and a surprising amount of diversity (Bellon *et al.* 2018). Maize is also the world's most productive crop (Ranum *et al.* 2014), and an important model system for genetics (Nannas and Dawe 2015).

Previous work in both maize and teosinte has demonstrated clear population structure at both regional (Pyhäjärvi *et al.* 2013) and fine (Van Heerwaarden *et al.* 2010) scales, and population genetic and common garden studies in both subspecies have shown clear signatures of populations being adapted to their ecological conditions. In maize this includes local adaptation to high elevation (Fustier *et al.* 2019; Gates *et al.* 2019), phosphorous (Rodriguez-Zapata *et al.* 2021), temperature (Butler and Huybers 2013), and day length (Swarts *et al.* 2017). Similarly, studies of teosinte have documented local adaptation based on features such as the differential patterns of microbial community recruitment (O'Brien *et al.* 2019), elevation (Fustier *et al.* 2019, 2017), and temperature and phosphorous (Aguirre-Liguori *et al.* 2019).

Studying local adaptation of maize and teosinte across the same geographic locations presents opportunities to disentangle multiple processes that interact with adaptation. For example, the effect of the domestication process in maize populations and their ongoing interaction and dependence on humans has created changes in the timing and types of selection imposed across all populations, as well as changes in demography (Wright *et al.* 2005). Based on population structure and differences in the abiotic environment among populations, we anticipated that local adaptation would have a small geographic scale. We predicted that sweeps would be exclusive to individual populations, and that adaptations shared between subspecies would be limited to sympatric pairs of populations growing in similar environments and with ample opportunity for genetic exchange. Because of domestication and the ongoing migration facilitated by humans, we expected that maize would show more shared adaptations, leading to a relatively larger geographic scale. Contrary to our predictions, our results suggest adaptations are often shared across two or more populations, and are commonly between maize and teosinte, rather than being exclusive to individual populations. We also found that migration and standing variation have played an important role as a source of beneficial alleles, including many that are shared across the two subspecies.

Results

We sampled teosinte (*Zea mays* subsp. *parviglumis*) individuals from six locations across its native range, along with a nearby (sympatric) population of traditionally cultivated open-pollinated maize (commonly referred to as landraces) at five of these locations (Figure 1C). We sampled ten individuals from each population for each subspecies, with the exception of the Palmar Chico populations, where we took advantage of 55 and 50 individuals previously sampled for maize and teosinte, respectively (Table S1, Supplement I, Chen *et al.* 2020). In most cases, results for both Palmar Chico populations were downsampled to ten randomly selected individuals to facilitate comparisons to the other populations. However, we used a second random sample to assess false positives in our inference of selective sweeps (see Results and Methods)

To further evaluate patterns of local versus geographically widespread adaptation, we compared many results of our geographically well-defined populations to results found over the entire study range (referred to as "rangewide" throughout). Rangewide samples for each subspecies were constructed by randomly selecting one individual from each population. All 195 individuals were sequenced at 20 to 25x coverage and aligned to version 5 of the *Zea mays* B73 reference genome (Hufford *et al.* 2021; Portwood *et al.* 2019). Analyses were based on genotype likelihoods (Korneliussen *et al.* 2014) except in cases where called genotypes were required (see Methods).

Subspecies and populations are genetically distinct despite evidence of gene flow.

To assess the relationships among our sampled populations, we constructed a population-level phylogeny using treemix (Pickrell and Pritchard 2012 v.1.13). As anticipated from previous work (Buckler and Holtsford 1996; Hufford *et al.* 2012), we found clear divergence between two clades composed of maize and

teosinte populations, though the relationship among geographic locations differed between the subspecies clades (Figure 1B).

Within subspecies, populations were genetically distinct from one another. Using NGSadmix (Skotte *et al.* 2013), there was little evidence of admixture between populations of the same subspecies; only two of the sampled individuals revealed mixed population ancestry (Figure 1A and 1D).

Despite the clear phylogenetic separation between the two subspecies, there is evidence for gene flow between maize and teosinte populations. We conducted f4 tests using treemix (Pickrell and Pritchard 2012), and found that all populations showed some evidence of gene flow with various populations of the other subspecies, as measured by the high absolute Z-Scores of the f4 statistic. However, Z-Scores were sensitive to the specific combinations of non-focal populations included in each test (Table S2). More specifically, we found that elevated f4 tests almost always included the maize population from Crucero Lagunitas ($p < 2 \times 10^{-10}$), which was true whether or not the f4 test included its sympatric teosinte, and more generally, there is little evidence for increased geneflow between sympatric pairs (Figure 1E).

Populations vary in their diversity, demography, and history of inbreeding

We estimated pairwise nucleotide diversity (π) and Tajima's D in non-overlapping 100Kb windows along the genome in our sampled populations using ANGSD (Korneliussen *et al.* 2014). For all populations, π was in the range of 0.006 to 0.01, consistent with both previous Sanger (Wright *et al.* 2005) and short-read (Hufford *et al.* 2012) estimates for both subspecies. Variation in Tajima's D and π was greater among populations of teosinte than maize (Figure 2D, Table S2).

We independently estimated the demographic history for each population from their respective site frequency spectra using mushi (DeWitt *et al.* 2021 v0.2.0). All histories estimated a bottleneck that started approximately 10 thousand generations ago (assuming a mutation rate of 3×10^{-8} (Clark *et al.* 2005) (Figure 2E).

Teosinte is a primarily outcrossing grass (Hufford 2010), and regional maize farming practices promote outcrossing as well (Bellon *et al.* 2018). To validate our estimated demography and characterize the history of inbreeding in each population, we compared the empirical quantiles of homozygosity by descent (HBD) segments inferred using IBDseq (Browning and Browning 2013) to those simulated under the demography of each population. With the exception of the smallest HBD segments, which are more prone to inaccurate estimation, the simulated quantiles generally resemble the empirical quantiles (Figure 2D). This indicates that the inbreeding history of our population is adequately captured by the demography. However, consistent with previous studies of teosinte (Hufford 2010), we do see variation in the distributions of HDB among populations. For example, the size distribution of HBD segments in San Lorenzo and Los Guajes were consistently larger than those simulated from their demographies, particularly for the smallest segments. This likely reflects inbreeding caused by demographic changes, particularly those further in the past that were not detected in our demography inferences. These results are consistent with previous studies, that found evidence for historical inbreeding in teosinte, particularly in individuals sampled from San Lorenzo (Pyhäjärvi *et al.* 2013). Lastly, we estimated average pairwise inbreeding coefficients (F) using ngsRelate (Hanghøj *et al.* 2019). Although for some pairs of individuals F was as high as 0.37, the mean value and F was 0.017 ± 0.001 (SE) and 0.033 ± 0.001 (SE) for maize and teosinte (respectively), (Figure 2E), suggesting there has been relatively little inbreeding in either subspecies in the recent past.

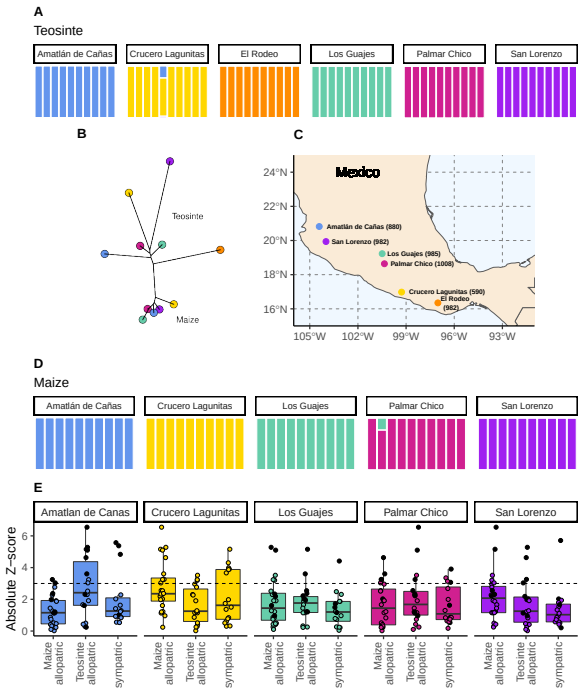

Figure 1 | The geographic distribution, population structure, and gene flow of maize and teosinte populations. (A and D) Admixture proportions among populations within subspecies. The dominant cluster in each population is colored by sampling location. (B) The unrooted tree of maize and teosinte populations. (C) Geographic sampling locations for the studied maize and teosinte populations. (E) F4 tests to quantify evidence of gene flow between the subspecies for allopatric and sympatric population pairs. Each point in (E) reports the absolute Z-Score for an f4 test, where a given focal population was partnered in a phylogenetic tree with another population of the same subspecies as a sister node, and two other populations from the other subspecies as a sister clade (see Methods for further details). Black points show f4 tests that included maize from Crucero Lagunitas, otherwise points are colored by focal population. Dotted line corresponds to our chosen significance threshold ($p = 0.001$).

Linked selection has a larger effect on diversity in maize

Census size is known to vary widely among maize and teosinte populations (Hufford 2010; Wilkes *et al.* 1967), yet genomewide pairwise sequence diversity (π) was relatively similar across our studied populations (Figure 2). Under strictly neutral models, census size differences should translate into differences in genetic diversity. Widespread selection at linked sites, however, can attenuate correlations between census size and diversity (Lewontin *et al.* 1974; Corbett-Detig *et al.* 2015; Buffalo 2021), and

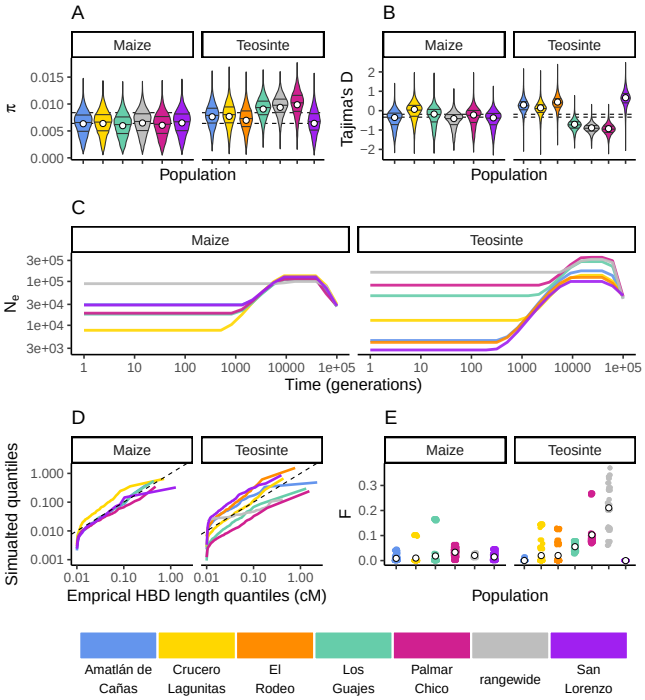

Figure 2 | Inbreeding, diversity, and demography. The distribution of π (A) and Tajima's D (B) calculated in 100Kb windows for maize and teosinte populations. Dashed lines show the median values for the two subspecies. Filled white points show the median values of each statistic generated from coalescent simulations under the demographic history inferred for each population. Colors for each population are as in Figure 1 and are shown at the bottom of the figure. (C) The inferred demography for each population. (D) The quantile of observed Homozygosity By Descent (HBD) lengths (cM) versus those simulated under each population demography. Dashed lines shows the 1:1 correspondence between the axes. (E) The distribution of inbreeding coefficients among pairs of individuals in each population. Filled white points are the average values for each population.

previous work has demonstrated differential impacts of linked selection on diversity in maize and teosinte (Beissinger *et al.* 2016).

In an effort to better understand the forces shaping variation among our studied populations, and to be able to better assess the reliability of our demographic inferences (which ignored linked selection), we fit a model of linked selection to predict π from estimates of recombination rate and the density of functional base pairs. As recombination rate increases, neutral variants are able to disassociate from nearby selected variants. Combined with gene annotation information, this signature allowed us to estimate what the population scaled mutation rate ($\theta = 4N_e\mu$) would be in the absence of linked selection (Coop and Ralph 2012; Corbett-Detig *et al.* 2015).

We used AIC (Wagenmakers and Farrell 2004) to select between different models that predicted neutral θ, finding that the full linked selection model, which includes background selection and hitchhiking parameters, was favored for all populations. In all populations, the estimated confidence intervals of neutral θ from the full model were non-overlapping with that of the intercept-only model which provides an estimate of the average genomewide value of π (Figure 3B). The two estimates of θ are relatively similar for all teosinte populations, while the difference was consistently larger in maize, suggesting linked selection has played a more important role in shaping maize diversity (Figure 3B). Similarly, the expected value of π approaches the estimated neutral diversity at lower recombination rate in teosinte (Figure 3A). Specifically, averaged across populations, π in teosinte reached 95% of the neutral θ estimate at a minimum recombination rate of 0.01 cM/Mb $\pm$0.0006 (SE) compared to 1.04 cM/Mb $\pm$0.05 (SE) in maize. Relatedly, in maize, approximately 92% $\pm$ 0.2% (SE) of the available genomic windows fell below the recombination rate minimum described above, compared to 12% $\pm$ 0.9% in teosinte. Previous inferences of linked selection in maize and teosinte found the reverse pattern, with a strong impact of linked selection in teosinte (Beissinger *et al.* 2016). One potential explanation for this discrepancy is that previous analyses did not directly estimate θ, but instead assumed that levels of π distant from coding sequences were reflective of the true neutral diversity. This approach would underestimate the effects of linked selection we observed in maize, where the majority of the genome has levels of π below θ. Overall, our results indicate that linked selection has played a larger role in shaping diversity for maize than teosinte populations.

Rangewide estimates of the proportion of mutations fixed by natural selection (α) are commensurate with that of individual populations

If populations are relatively isolated and adaptation occurs primarily via local selective sweeps, then we expect that most adaptive fixations will happen locally in individual populations rather than across the entire species range. If adaptation via sweeps is commonly restricted to individual populations, using a broad geographic sample to represent a population could underestimate the number of adaptive substitutions that occur (Gossmann *et al.* 2010). To test this, we estimated the proportion of mutations fixed by adaptive evolution (α) (Smith and Eyre-Walker 2002) across all of our populations and the rangewide samples for both subspecies. We estimated α jointly among all populations by fitting a non-linear mixed-effect model based on the asymptotic McDonald–Kreitman test (Messer and Petrov 2013). Across populations, α varied between 0.097 (teosinte San Lorenzo) and 0.282 (teosinte Palmar Chico), with more variation among teosinte populations (Figure 3C). In contrast to our expectations, rangewide estimates of α were commensurate with individual populations. We additionally evaluated estimates of α for specific mutation types, which has been shown to be lower at sites mutating from A/T to G/C, due to the effects of GC biased gene conversion in *Arabidopsis* (Hämälä and Tiffin 2020). While we do find some evidence that α predictions varied by mutation type (see Figure S3, Supplement IV), the patterns are the opposite of that found in *Arabidopsis*, perhaps because of the increased level of methylation in maize and the higher mutation rate at methylated cytosines. Even after accounting for differences among mutation types, rangewide values remained commensurate with that of the populations.

The proportion of mutations fixed by natural selection (α) is associated with neutral diversity in teosinte

The variable demographic histories of our populations may have affected the process of adaptive evolution. Under a neutral mode

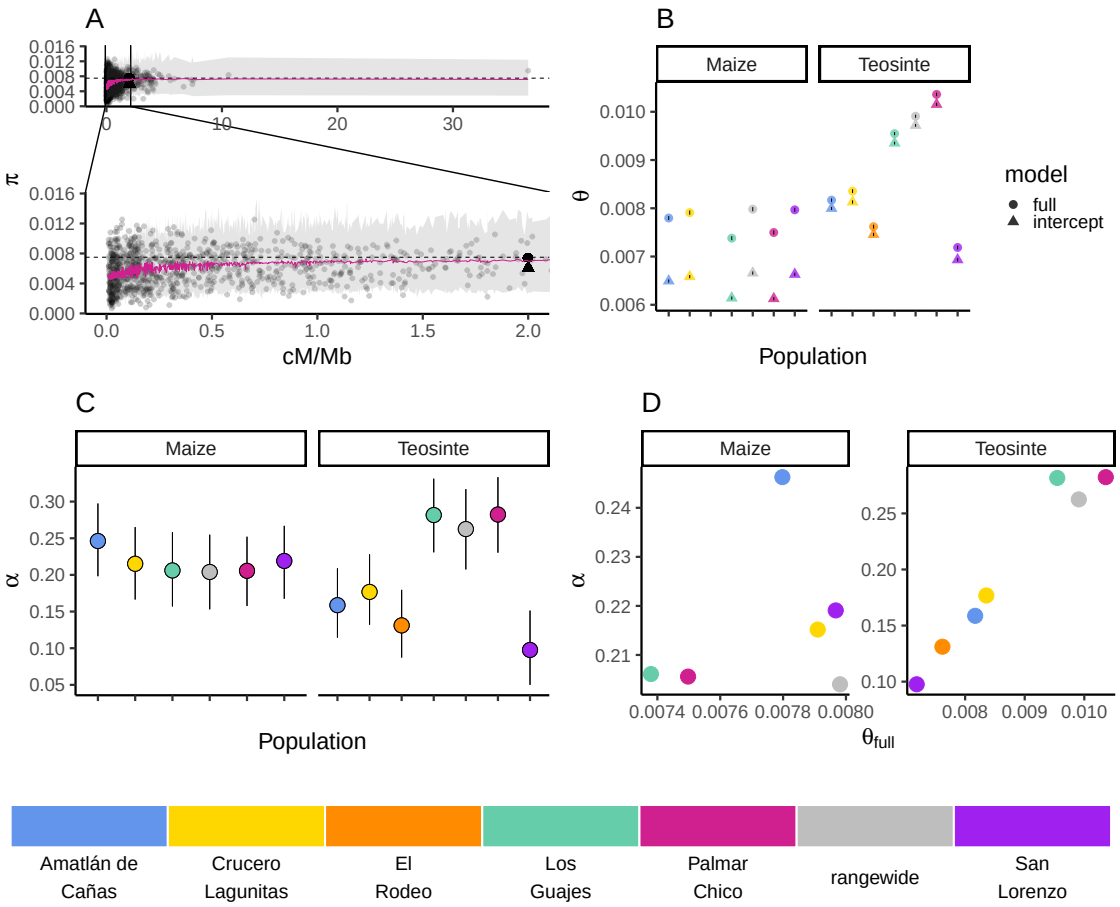

Figure 3 Linked selection and the proportion of mutations fixed by natural selection. (A) The relationship between recombination rate (cM/Mb) and diversity (π) measured in 10 Kb windows, overlaid with the full linked selection model fit (in purple; data and fit shown for maize from Palmar Chico only). Grey bands show the 95% prediction interval of the full model fit. The dotted line shows the maximum likelihood estimate of neutral θ. Bottom of (A) shows zoomed in region of higher density. The circle and triangle at the edge of the zoomed in plot show the maximum likelihood estimates of θ for the full and intercept models (see Methods), as do the points shown in panel B. (B) The maximum likelihood estimates of θ for the full and intercept models for each population, respectively. (C) Estimated values of the proportion of mutations fixed by natural selection (α). Vertical lines show the 95% credible interval. (D) The relationship between (α) and the maximum likelihood estimate of θ from the full linked selection model for each population.

of evolution, population size and selection have no impact on the rate of substitutions, which are predicted to be equal to the mutation rate. When adaptive mutations are introduced, theory predicts that the rate of substitutions depends on the mutation rate, effective population size, and the strength of selection (Gillespie 2004). It follows that populations with larger long-term effective populations sizes should have a higher proportion of substitutions that are adaptive.

Motivated by theory, we used simulations to confirm that α increases with population size, and that the relationship is robust to variation in recent demographic changes and the distribution of fitness effects (table S3 and figure S2). We then investigated the empirical relationship between θ and α in our populations by testing for an association between the two separately for each subspecies. We used the mean α predictions and associated them with the values of neutral θ estimated from the full linked selection model. We found that θ and α are strongly correlated for teosinte populations ($r = 0.97$, $p < 3.6 \times 10^{-9}$), but not for maize ($r = 0.09$, $p = 0.77$) (Figure 3).

Teosinte populations have a higher proportion of private sweeps

Our inferences of α are based on substitutions at non-synonymous sites. The functional space for selection to act on occurs over many other parts of the genome besides protein coding bases, especially for large repetitive plant genomes (Mei *et al.* 2018). To identify signatures of adaptation occurring anywhere in the genome, we used RAiSD (Alachiotis and Pavlidis 2018) to identify putative selective sweeps in each population, where sweeps were selected using an outlier cutoff defined by the 99.9% percentile of the μ summary statistic for data generated using coalescent simulations under each population's estimated demography (see Methods). Simulations suggest this approach has high accuracy and power compared to alternative

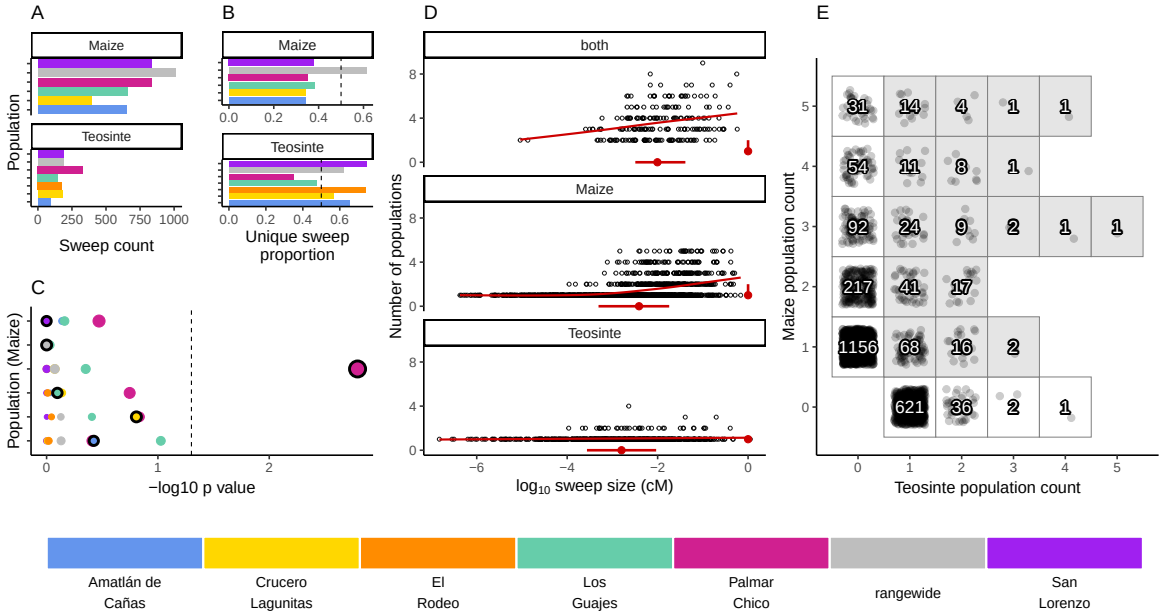

Figure 4 The distribution of shared and private selective sweeps. (A) The total number of sweeps inferred in each population. (B) The proportion of sweeps that are unique to each population. Dotted line drawn at 50%. (C) The negative log10 p values for hypergeometric tests to identify maize-teosinte population pairs that shared more sweeps than expected by chance (see Methods). P values were adjusted for multiple tests using the Benjamini and Yekutieli method. Populations along the y axis are maize (order matches the legend below, with Amatlán de Cañas on bottom), while the point color designates the teosinte population each maize population was paired with. Points with black outlined highlight the sympatric population comparisons. Point size is scaled by the number of shared sweeps identified in each pair. The dotted line indicates our chosen significance level ($p = 0.05$). (D) The relationship between genetic map length of sweep regions and the number of populations each sweep region occurred in, split for sweeps shared in two or more populations of both subspecies, exclusive to maize, or exclusive to teosinte. The black line shows a loess smooth fit. Grey points and bars at the margin of each subplot show the 50th, 25th and 75th percentiles of sweep sizes, respectively. (E) Counts of shared and unique sweeps broken down by how many maize and teosinte populations they occurred in. Grey boxes show sweeps shared across the two subspecies.

methods over a broad range of demographic and sweep scenarios (Alachiotis and Pavlidis 2018). To further assess the accuracy of our sweep inferences, we compared the overlap in sweep regions from a second random sample from both maize and teosinte from Palmar Chico. We estimated a false positive rate to be 0.6 and 0.5 for maize and teosinte, respectively, suggesting many of the putative sweep regions in other populations are also false positives, and that the total number of sweeps we have identified is likely an overestimate. Approximately 10% of all base pairs fell within a sweep region found in one or more of the sampled populations, although the true proportion is likely lower based on our estimated false positive rate. The density of coding sequence was lower in sweep regions (0.029 coding sequence base pairs/sweep base pairs) than the genome-wide gene density (0.042 coding sequence base pairs/total base pairs).

We used the inferred sweep regions to assess the degree to which adaptation is shared or locally restricted using the sweep regions we identified. We determined how many sweep regions were exclusive to one population (private), along with the number and size of overlapping sweep regions shared across two or more populations within and between the two subspecies. Overall, sharing was common, though fewer sweeps were exclusively shared between teosinte populations (Figure 4A). In teosinte populations, 59% of sweeps were private, which was significantly greater than the 40% found in maize (binomial glm, $p = 2.63 \times 10^{-7}$; Figure 4B). We saw a general pattern that larger sweep regions tended to be shared between more populations, though the pattern was negligible for sweeps exclusive to teosinte (Figure 4D). Additionally, 31 out of 1550 sweep regions were exclusive to and shared by all maize populations, compared to 0 out of 660 for teosinte.

Sympatric population pairs do not share more sweeps

If local adaptation favors certain alleles in a given environment, we might expect to see increased sharing of sweeps between sympatric populations of maize and teosinte. To look for evidence of such higher sharing, we used a hypergeometric test based on the number of sweeps in each population, and the number of shared sweeps between population pairs, which allowed us to test if sympatric population pairs tended to have more sharing than expected by chance. In conducting this test, we incorporated the estimated false positive rate (see Methods). Sympatric pairs did not tend to have a lower p value than allopatric pairs, and only the population pair in Palmar Chico showed more sharing than expected by chance (Figure 4). We additionally found that, despite evidence that sweeps are commonly shared between maize and teosinte (Figure 4C), there were zero sweeps exclusive to sympatric pairs; sweeps that were shared between sympatric pairs always included at least one other allopatric population.

Convergent adaptation from migration is common among maize and teosinte populations

In instances when two or more populations shared a sweep region, we used rdmc (Lee and Coop 2017; Tittes 2020) to infer the most likely mode of convergence. We classified sweeps based on which composite log-likelihood model was greatest out of four possible models of convergence (independent mutations, migration, neutral, and standing variation). Of the 1734 sweeps that were shared by two or more populations, there were 17, 663, 627, and 427 sweeps inferred to be convergent via independent mutations, migration, neutral, and standing variation, respectively (Figure 5C). While the high proportion of neutral models inferred by rdmc is consistent with a relatively high false positive rate, confirmation of non-neutral patterns of diversity from this second approach increases our confidence that regions identified for other modes indeed represent sweeps. The strength of support (measured as the composite likelihood score of the best model relative to the next best) varied among sweeps and modes of convergence, but in general a single model tended to be clearly favored among the alternatives (Figure 5A). Selection coefficients for sweeps varied among modes, with convergence via migration having the highest average estimate (Figure 5B). When migration was the mode of convergence and sweeps were shared by both subspecies, teosinte Palmar Chico and maize from Crucero Lagunitas were the most frequent source populations (Figure 5D). In convergence models with migration, we considered only a low and high migration rate, set to 0.001 and 0.1, respectively. The most likely migration rate varied across sweeps, but the migration rate was high for a majority of sweeps shared by the two subspecies and sweeps exclusive to maize. In contrast, the low migration rate was always most likely for sweeps exclusive to teosinte (Figure 5F). Together, these findings indicate that many alleles are adaptive in the genomic background of both maize and teosinte, and that adaptive alleles are commonly shared between the two subspecies.

Discussion

Local adaptation occurs at intermediate scales

Our findings overall suggest that adaptation in maize and teosinte does occur locally, but frequently at scales larger than individual populations. Gossmann *et al.* (2010) hypothesized that population structure within a species could limit the fixation of adaptive alleles across a species range, causing a reduction in the proportion of mutations fixed by positive selection (α). Based on this hypothesis and the strong population structure we observed (Figure 1), we expected that rangewide samples would have a smaller estimates of α. Instead, α for the rangewide samples of both maize and teosinte were commensurate with that of individual populations (Figure 3), a pattern that persisted even when we considered α estimated from several different mutation types (Figure S3). This is inconsistent with the patterns we would expect fine-scale local adaptation to generate, where adaptive substitutions for a given population should not be shared by other populations experiencing their own distinct local selective pressures.

We found a similar pattern from our analysis of shared versus unique selective sweeps, which were more often shared by at least one other allopatric population. Similar to our predictions for α, we expected that local adaptation would lead most sweeps to be exclusive to individual populations. Instead, the average proportion of sweeps exclusive to a single population was low to moderate for maize and teosinte populations, respectively (Figure 4). We also expected that maize and teosinte populations growing in close proximity would share similar local selective pressures and would therefore share more signatures of adaptation. However, only maize and teosinte sampled from Palmar Chico showed evidence of sharing more sweeps than would be expected by chance, and overall sympatric pairs did not show increased sharing of selective sweep regions compared to allopatric pairs (Figure 4). The regional scale of local adaptation is consistent with patterns seen in maize adaptation to the highlands (Calfee *et al.* 2021), where sympatric maize and teosinte populations show little evidence of adaptive gene flow, and adaptive teosinte introgression appears widespread among highland maize. To our surprise, private sweeps made up a higher proportion of sweeps for the rangewide sample of maize. One possible explanation is that rangewide samples are enriched for older sweeps that are shared across all populations but difficult to detect in any single sample. Consistent with this, we find evidence that the sweep statistic μ for individual maize populations was elevated in the sweep regions exclusive to the rangewide sample, but did not reach the cutoff to be considered outliers (Figure S6, 22).

There are a number of considerations to make in the interpretation of our results. The two methods we used to identify signatures of adaptation, estimating α and identifying signatures of selective sweeps, both depend on the fixation of beneficial mutations. For the moderate population sizes and selection coefficients observed here, fixation of new beneficial mutations takes a considerable amount of time, on the order of $4log(2N)/s$ generations (Charlesworth 2020) of thousands of years. Compared to the sojourn time of adaptive mutations, our populations may have occupied their current locations for relatively few generations. As a result, the selective sweeps underlying local adaptation to the selective pressures that populations currently face are more likely to be incomplete, so may be more difficult to detect (Xue *et al.* 2021; Pritchard *et al.* 2010). Likewise, the adaptive sweeps that have completed — and are more likely to be detected — may have been under selection in ancestral populations that occupied different environments than that of the sampled individuals. Another complication in detecting local adaptation relates to the size and complexity of plant genomes. Large genomes may lead to more soft sweeps, where no single mutation driving adaptive evolution would fix (Mei *et al.* 2018). Like incomplete sweeps, soft sweeps are harder to identify (Schrider and Kern 2016; Pritchard *et al.* 2010), which could obscure the signatures of local adaptation. Even if our populations have occupied their current locations for a sufficient duration for local adaptation to occur, the completion of selective sweeps may be hindered by changes and fluctuations in the local biotic and abiotic conditions. Relatively rapid change in local conditions could result in fluctuating selection, such that most alleles do not remain beneficial for long enough to become fixed (Rudman *et al.* 2021).

While our focus has been on the trajectory of individual beneficial alleles, the genetivc basis of many adaptive traits may be highly polygenic. Allele frequency changes underlying polygenic adaptation are more subtle than those assumed under selective sweeps, making them harder to detect (Pritchard *et al.* 2010). Evaluating local adaptation in maize and other systems will be facilitated by studying the contribution of polygenic adaptation to the evolution of complex traits. However, if adaptation across our studied populations were strictly polygenic,

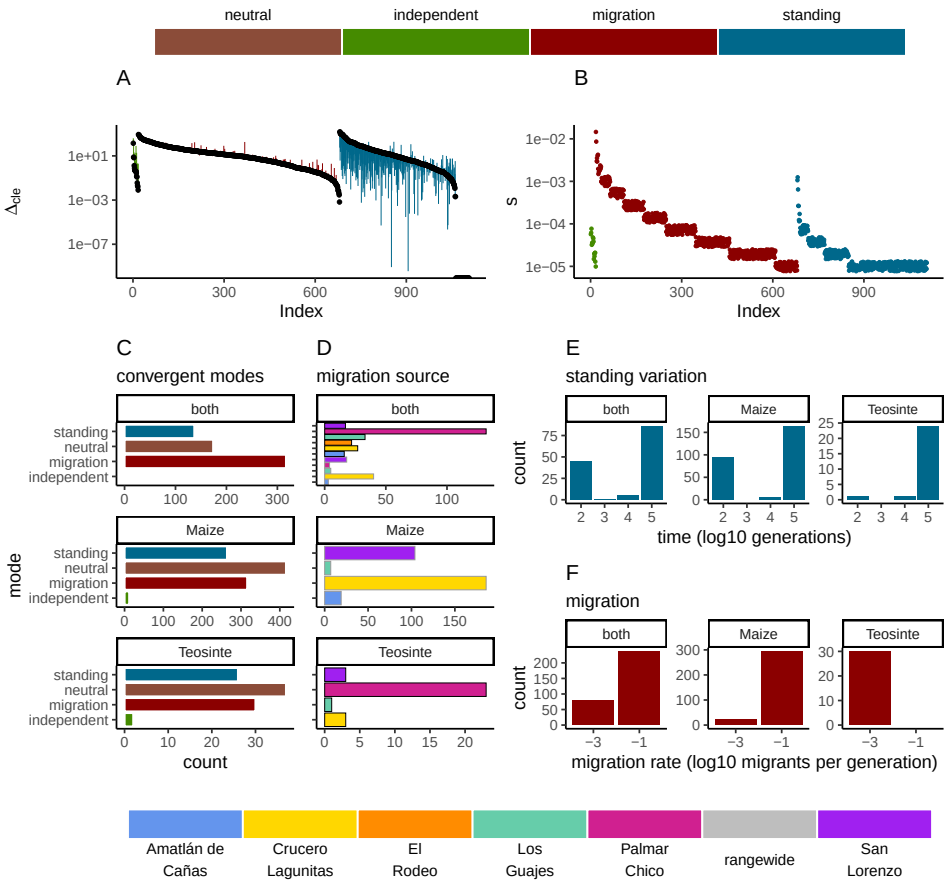

Figure 5 Modes of convergent adaptation and affiliated parameters for shared selective sweeps. (A) The difference in composite likelihood scores for the best supported mode of convergent adaptation (colors in top legend) compared to next best mode (black points), and best mode compared to the neutral model (other end of each line segment above or below black point). (B) Selection coefficients colored by the most likely mode of convergent adaptation. (C) Number of shared sweeps for both subspecies that were inferred to be from each convergent adaptation mode. (D) The most likely source population for shared sweeps that converged via migration. Bars are colored by population (bottom legend) and are outlined in black for teosinte and grey for maize. (E) Observed frequency of the inferred time in generations that each selected allele persisted prior to selection for models of convergent adaptation via standing variation. (F) Observed frequency of each inferred migration rate value for models of convergent adaptation via migration. Panels C, D, E and F are partitioned by which subspecies shared the sweep.

and especially if it were acting on alleles with small effect sizes, we would expect to find little or no shared sweeps. The fact that we find many instances of sharing across populations supports that a non-trival amount of local adaptation is occurring via selective sweeps, or through polygenic adaptation acting on a few loci with large effects that leave a signature similar to that of a selective sweep.

Differences in diversity and demography influence adaptation in maize and teosinte

While our results were generally similar between the two subspecies and among the sampled populations, there are several important differences. The most obvious difference between the subspecies is the ongoing interaction and dependence of maize on humans via domestication and farming. Compared to teosinte, maize had lower average genomewide estimates of diversity (Figure 2A), which persisted after accounting for the effects of linked selection (Figure 3B). These differences are consistent with the previously discovered pattern that diversity tends to be lower in crops compared to their wild relatives (Doebley 1989; Hufford *et al.* 2012), a pattern putatively driven by domestication bottlenecks (Eyre-Walker *et al.* 1998). In line with this argument, the few teosinte populations with lower diversity than those in maize (El Rodeo and San Lorenzo) were inferred to have the most substantial bottlenecks and historical inbreeding (Figure 2). More generally, we found that π and Tajima's D were more variable among teosinte populations, indicative of differences in their demographic histories.

Our demographic inferences suggest that all populations had signatures of a bottleneck, the timing of which coincides with the beginning of maize domestication $\approx 9,000$ years ago (Piperno *et al.* 2009). The severity of the bottleneck varied considerably across populations, particularly for teosinte. While finding a bottleneck in the maize populations is consistent with domes-

tication, it is less clear why we found a consistent bottleneck for the teosinte populations at approximately the same time. One possibility is that the teosinte bottlenecks reflect land use change induced by human colonization. For example, evidence from Mesoamerican phytolith records in lake sediment show evidence of anthropogenic burning as early as 11K years B.P. (Piperno 1991). The establishment and spread of human populations over the subsequent millenia would require ever increasing area for farming, dwellings, transportation, and trade (Haines *et al.* 2000). Such land use changes would likely encroach on the habitat available for teosinte and drive species-wide census size declines. Given the success of maize breeding and domestication, we anticipated a recent expansion for maize populations as previously seen (Beissinger *et al.* 2016; Wang *et al.* 2017). However, with only 10 individuals per population, recent expansions will be difficult to detect (Keinan and Clark 2012; Coventry *et al.* 2010). The demography of the rangewide samples for both subspecies showed little evidence of the bottleneck inferred in individual populations, likely due to the reduced sampled size (5 and 6 individuals) for the rangewide data. We additionally used strong regularization penalties to avoid overfitting (see Methods), which limits the detection of rapid and dramatic changes in population size. The near-constant size of the rangewide samples and the lack of recent expansion in maize are both likely influenced by this modeling choice. Lastly, given the strong effects of linked selection inferred in maize, our demographic models may have underestimated their effective population sizes. However, all populations independently converged to similar values in the oldest generation times, around the time when we would expect the ancestral lineages would have coalesced (Figure 2C). This suggests any biases in the estimated population sizes that are specific to maize are occurring in the more recent past. Similar to the arguments above, this provides another reason that recent population sizes in Maize may be underestimated.

Theory and simulations, including those presented here, support that the rate of fixation of new beneficial mutations should increase with the population size (N_e) and consequently the population-scaled mutation rate ($\theta = 4N_e\mu$) (Gillespie 2004, Figures 3D and Supplement III). However, previous research in plants found little evidence for positive estimates of α, and an apparent lack of correlation between inferred population size and α for new non-synonymous mutations (Gossmann *et al.* 2010). The discrepancy between those results and our own, in regards to the overall estimates of α, are likely explained by the use of different methods and the sampling design. Our simulations suggest the relationship between α and θ is positive but non-linear. The rate of increase for α declines at larger values of θ (Figure S2) This suggests finding an association between θ and α may be sensitive (in theory and practice) on the parameter space under consideration.

We found a significant correlation between neutral θ and α, but only for teosinte populations (Figure 3C). While we consider this finding preliminary based on the small number of populations in the sample, the positive association found in teosinte populations is consistent with our simulations and some theory cited above that suggest larger populations increase the rate of adaptive fixation. However, it has also been suggested that the relationship between effective population size (N_e) and α is driven by increased efficiency of purifying selection to reduce the fixation of nearly-neutral substitutions, rather than an increase in the fixation rate of beneficial mutations *per se* (Galtier 2016; Hämälä and Tiffin 2020). We cannot confidently determine whether the association is driven by the increased efficiency of larger populations to purge weakly deleterious mutations or to fix adaptive ones. However, these are not mutually exclusive explanations, and both may contribute to differing degrees in actual populations. The lack of correlation in maize is likely driven by the recent divergence between populations and the strong recent selection during domestication that is shared by all the populations. As a result, populations have not had sufficient time for differences in population size to impact patterns of adaptive fixation. As recent changes in environment and population size are likely important aspects of the biology of many domesticates, we predict that the observed lack of correlation between θ and α in maize may be a common pattern across many taxa. However, another important caveat concerning these associations is the lack of independence among lineages, particularly for maize populations, which have diverged more recently. Further, studies with larger sample of populations may correct for non-independence when testing the associations between α and θ, which would bolster the validity of any associations (or lack there) found.

Although the asymptotic MK method we employed has been shown to provide reliable estimates of α when fixations are due to strong beneficial mutations (Messer and Petrov 2013), it does not account for the influence of background selection and the rate of fixation of weakly beneficial mutations (Uricchio *et al.* 2019). Our analyses suggests linked selection is acting similarly across populations within subspecies (Figure 3B). As such, while we may be systematically underestimating α, the bias is unlikely to change our general conclusions.

Another consideration is that our estimates of α contrasted non-synonymous and synonymous sites only. We observed a lower density of coding sequence in sweep regions, suggesting selective sweeps are more often acting on non-coding regulatory regions rather than modifying protein coding sequences directly. This result is consistent with the proposed importance of non-coding sequences for adaptation in species with large genomes (Mei *et al.* 2018) and highlights the importance of studying signatures of adaptation using multiple methods and regions of the genome in order to construct a comprehensive history.

Differences in adaptation between maize and teosinte, and among populations, were apparent based on differences in the patterns of selective sweeps. Maize had a higher proportion of selective sweeps shared with at least one other population (Figure S1). The greater number of shared sweeps in maize populations is likely the result of their recent shared selective history during the process of domestication, resulting in a set of phenotypes common to all maize (Stitzer and Ross-Ibarra 2018). In comparison, the higher proportion of unique sweeps in teosinte suggests local adaptation has played more of a role in shaping their recent evolutionary history. Teosinte grows untended, and did not undergo domestication, leaving more opportunity for divergence and local selection pressures to accumulate differences among populations. This is reflected in the inferred population history, which had longer terminal branch lengths for teosinte (Figure 1B), suggesting there is increased genetic isolation among teosinte populations due to longer divergence times, reduced gene flow, or both.

Convergent adaptation is ubiquitous

We found convergent adaptation to be common among populations and subspecies (Figures 4 and 5). The frequency of

convergence further suggests there are a large number of mutations that are beneficial in more than one population, even when placed in the different genomic backgrounds of the two subspecies. Our approach allowed us to distinguish between multiple potential modes of convergence, including a neutral model that models allele frequency covariance by drift alone (Lee and Coop 2017). The distribution of most likely selection coefficients of the inferred beneficial alleles suggests the strength of selection is moderate to strong, though this estimate is likely biased as strong positive selection will be easier to detect. Convergence via independent mutations was by far the least frequent mode. This is consistent with previous analyses of domestication (Hufford *et al.* 2012) and adaptation (Wang *et al.* 2020) in maize, and unsurprising given evidence for ongoing gene flow (Figure 1E), the relatively short evolutionary time scales, and the low probability that even strongly selected new mutations can overcome drift multiple times independently. When convergent sweeps that occurred via standing variation within maize, or for those shared between maize and teosinte, the distribution of generation times that the selected variant was standing before the onset of selection tended to be bimodal, with both long and short standing times. Whereas sweeps exclusive to teosinte were consistently inferred to be standing variation for more generations (Figure 5E). Sweeps that occurred via standing variation and shared between subspecies were often found in only a subset of maize populations. Many of these sweeps likely reflect the presence of structure in ancestral populations, suggesting different alleles beneficial to maize were likely derived from more than one teosinte population.

The most common mode of convergent adaptation was via migration, and frequently occurred between geographically desperate populations (Figures S4). This included a relatively large number of shared sweeps via migration between maize and teosinte (Figures 5 and S5). There is ample evidence that maize and teosinte are capable of hybridizing (Wilkes *et al.* 1967; Ellstrand *et al.* 2007; Ross-Ibarra *et al.* 2009). Gene flow between geographically disparate populations has been inferred from patterns of introgression between maize and *Zea mays mexicana*) (Calfee *et al.* 2021). Further, convergence via migration between geographically disparate maize populations has been inferred during adaptation to high elevations (Wang *et al.* 2020).

Our findings on convergence via migration point to an intriguing hypothesis, namely, that some non-trivial number of alleles that are beneficial to teosinte may have originally arisen in another teosinte population and moved between populations via gene flow with maize, an idea suggested by Ross-Ibarra *et al.* (2009) based on allele sharing at a small set of loci. There are several aspects of our results consistent with this hypothesis. Firstly, we found relatively few shared sweeps exclusive to teosinte populations (Figure 5C), which is what we would expect if maize populations facilitate the movement of beneficial teosinte alleles. However, it is important to note that there are fewer sweeps exclusive to teosinte for all modes of convergence, not just via migration, so this alone is not sufficient evidence. Secondly, sweeps shared via migration that were exclusive to teosinte were always inferred to be the lower of the two migration rates (1×10^{-3}) (Figure 5F), where those shared by both subspecies or exclusive to maize were primarily inferred to be the higher migration rate (1×10^{-1}). This indicates alleles that are only beneficial in teosinte move between populations more slowly. Despite the patterns of population structure in both subspecies, there is evidence of gene flow based on f4 tests. In particular, f4 tests that included maize from Crucero Lagunitas were consistently elevated across both subspecies (Figure 1E and S1). This is in accordance with our finding that Crucero Lagunitas maize was the second most commonly inferred source population of migration sweeps that were shared between maize and teosinte (Figure 5D). Together, these results suggests that geographically widespread varieties of maize such as Celaya (Crucero Lagunitas) (Orozco-Ramírez *et al.* 2017) may have played a prominent role as a source of and/or transport for beneficial alleles among maize and teosinte populations. However, teosinte populations were more often the source of beneficial alleles for sweeps shared between both subspecies. In particular, the teosinte Palmar Chico was the most likely source population for the vast majority of sweeps shared between maize and teosinte as well as for sweeps exclusive to teosinte (Figure 5F). Palmar Chico is in the balsas region, near the putative site of maize domestication (Piperno *et al.* 2009), and our demographic inferences suggests Palmar Chico has maintained the largest effective population size of all the populations (Figure 2), meaning it is capable of generating the most mutations for selection to act on. As such, teosinte Palmar Chico and other populations in the Balsas region may have also been a particularly important source of adaptive variation for both teosinte and maize.

Materials and Methods

Samples and whole genome re-sequencing

We sampled seeds from five populations of *Zea mays* ssp. *parviglumis*, and four populations of *Zea mays* ssp. *mays* from plants growing across the species' native range. We additionally included populations of *parviglumis* and maize from Palmar Chico were previously analyzed and reported in (Chen *et al.* 2020) for a total of 6 and 5 populations of maize and teosinte (See Supplement I for accession IDs and further sample details). All maize and teosinte populations from each named location were less than 1km from one another, with the exception of Crucero Lagunitas, which were separated by approximately 18km.

DNA extraction for teosinte followed (Chen *et al.* 2020). Genomic DNA for landraces was extracted from leaf tissue using the E.Z.N.A.® Plant DNA Kit (Omega Biotek), following manufacturer's instructions. DNA was quantified using Qubit (Life Technologies). 1ug of DNA per individual was fragmented using a bioruptor (Diagenode) with 30 seconds on/off cycles.

DNA fragments were then prepared for Illumina sequencing. First, DNA fragments were repaired with the End-Repair enzyme mix (New England Biolabs). A deoxyadenosine triphosphate was added at each 3'end with the Klenow fragment (New England Biolabs). Illumina Truseq adapters (Affymetrix) were then added with the Quick ligase kit (New England Biolabs). Between each enzymatic step, DNA was washed with sera-mags speed beads (Fisher Scientific). The libraries were sequenced with an average coverage of 20 to 25x PE150 on the Xten at Novogene (Sacramento, USA).

We additionally grew one individual of *Zea diploperennis* from the UC Davis Botanical Conservatory as an outgroup. DNA for *Zea diploperennis* was extracted and libraries prepared as above, and then sequenced to 60X coverage using PE250 on 3 lanes of Illumina 2000 rapid run (UC Davis Genome Center, Davis, USA).

Sequencing reads have been deposited in the NCBI Sequence Read Archive under project number (to be submitted).

Sequencing and variant identification

All paired-end reads were aligned to version 5 of the maize B73 reference genome (Hufford *et al.* 2021) using bwa-mem (v0.7.17) (Li 2013). Default options were used for mapping except -M to enable marking short hits as secondary, -R for providing the read group, and -K 10000000, for processing 10 Mb input in each batch. Sentieon (v201808.01) (Freed *et al.* 2017) was used to process the alignments to remove duplicates (option –algo Dedup) and to calculate various alignment metrics (GC bias, MQ value distribution, mean quality by cycle, and insert size metrics) to ensure proper mapping of the reads.

All downstream analyses were based on genotype likelihoods estimated with *ANGSD* (v0.934) (Korneliussen *et al.* 2014) using the following command line flags and filters:

```
-GL 1
-P 5
-uniqueOnly 1
-remove_bads 1
-only_proper_pairs 1
-trim 0
-C 50
-minMapQ 30
-minQ 30
```

Genetic Diversity

We estimated per base nucleotide diversity (π) and Tajima's D (D) in non-overlapping 1kb, 10kb and 100kb windows with *thetaStat* utility from *ANGSD*, though estimates did not substantively differ between window sizes. To estimate π and the unfolded site frequency spectra for each population, we polarized alleles as ancestral and derived using short-read sequence data for *Zea luxurians* and *Zea diploperennis* as outgroups. *Zea luxurians* sequence from Tenaillon *et al.* (2011) was downloaded from The NCBI Sequence Read Archive (study SRR088692). We used the alignments from the two species to make minor allele frequency (MAF) files using *ANGSD*. We used the MAF files to construct a table of genotypes found at each locus. Sites with minor allele frequency estimates greater than 0.001 were treated as heterozygous. Sites that were homozygous in both species were imputed onto the maize v5 reference and assumed to be the ancestral allele. As there were substantially more called bases in *Zea luxurians* than in *Zea diploperennis*, we also assumed sites that were homozygous in *luxurians* and missing in *diploperennis* were ancestral, but excluded sites that were missing from *luxurians*. Sites that were classified as heterozygous were treated as missing and imputed onto the maize reference as 'N'.

Population Structure and Introgression

We used *ngsadmix* (Skotte *et al.* 2013) to assess population structure within subspecies. To do so we used a SNP-calling procedure in *ANGSD* with the same filters as listed above, along with a SNP p value cutoff of 1×10^{-6}. We looked for evidence of gene flow between subspecies using f4 statistics and Z-scores calculated with blocked jackknifing, implemented using treemix (Pickrell and Pritchard 2012). Trees for f4 tests were always of the form $(Maize_X, Maize_X; Teosinte_focal, Teosinte_X)$; each unique combination of populations was considered to be the "$focal$" and "$_X$" positions of the tree. We considered any tree with a Z-score greater than or equal to 3 significant, indicating a departure from the allele frequency differences expected if population history matched the hypothesized tree. We assessed Z-scores separately based on whether the focal population was maize or teosinte, and for trees that include the sympatric pair of the focal population.

Demographic and Inbreeding History

We inferred each population's demography using a single unfolded site frequency spectrum with mushi (v0.2.0) (DeWitt *et al.* 2021). In efforts to reduce overfitting given our modest samples sizes, we increased the regularization penalty parameters to alpha_tv=1e4, alpha_spline=1e4, and alpha_ridge = 1e-1.

We assessed homozygosity by descent (HBD) in each population using IBDseq (v2.0) (Browning and Browning 2013). We compared empirical results to simulations in msprime (Kelleher *et al.* 2016) using each population's inferred demographic history. We performed ten replicates of each of these simulations. Replicates were similar across all populations; only one replicate was chosen at random for visual clarity.

We estimated recent inbreeding using ngsrelate (Hanghøj *et al.* 2019) with default parameters. Input files were generated using *ANGSD* with the same filters as listed above, along with a SNP p value cutoff and maf filter of

```
-SNP_pval 1e-6
-minmaf 0.05,
```

respectively.

Linked Selection

Following an approach similar to Corbett-Detig *et al.* (2015), we modeled the effects of linked selection (background selection and hitchhiking) on observed values of π across the genome in 1kb windows.

The full model that jointly modeled both hitchhiking and background selections was

$$\pi_i = \frac{\theta_{neutral}}{(1/e^{-G_i} + \alpha * fd_i/R_{bp_i})}. \tag{1}$$

Here, α is the population-scaled rate of selective sweeps (Coop and Ralph 2012), fd_i is the proportion of functional bases in window i, and R_{bp_i} is the mean recombination rate per base pair in window i. Lastly, G_i models the effects of background selection and is defined as,

$$G_i = \Sigma_k \frac{U * fd_k * sh}{2 * (sh + P(|M_i - M_k|) * (sh + P(|M_i - M_k|)} \tag{2}$$

where, U is the deleterious mutation rate, s and h are the selection and dominance coefficients for deleterious mutations (respectively), P is an index of panmixis, and M_i and M_k represent the Morgan position of the focal locus and all other loci along the chromosome, respectively (Hudson and Kaplan 1995). The proportion of functional bases fd_i, was estimated using the number of protein coding sequences in a window based on the most gene annotation of the maize v5 reference genome (Hufford *et al.* 2021). The position of loci in Morgans was approximated using the *approx* function in R based on a previously generated genetic map for maize that was lifted on to the v5 reference genome (Ogut *et al.* 2015). We treated the parameters s and U as fixed, but refit the model over a grid of possible values for each. Specifically, we allowed s to vary over 30 evenly spaced values between 1×10^{-8} and 1×10^{-1}, U took on values 1×10^{-7} or 3×10^{-8}, while P and h were always fixed at 1 and 0.5, respectively.

We compared the full linked selection model (equation 1) to three simpler models: one that does not account for hitchhiking

$$\pi = \frac{\theta_{neutral}}{(1/e^{-G_i})}, \tag{3}$$

one that does not include background selection,

$$\pi = \frac{\theta_{neutral}}{1 + \alpha * fd_i/R_{bp_i})} \tag{4}$$

and finally a simple intercept model that estimates neutral θ

$$\pi = \theta_{neutral}. \tag{5}$$

The same refitting procedure was used for all models that included U, s, and h. We fit all four equations separately for each population using a Maximum Likelihood framework. Since all values of θ are bounded between (0,1), we assumed observations were generated from a beta distribution following

$$\mathbf{L} = \prod_i beta(\pi_i * \sigma, (1 - \pi_i) * \sigma). \tag{6}$$

We calculated the standard errors for parameters from the Hessian matrix ($\mathbf{H}$) produced by the *optim* function as

$$CI_{parameter} = parameter \pm 1.96 * \sqrt{diag(\mathbf{H}^{-1})}. \tag{7}$$

We calculated AIC, ΔAIC and $AIC\ weights$ (Wagenmakers and Farrell 2004) and used them to compare among the alternative models for each population.

Estimating the Rate of Positive Selection, α

We modeled the rate of positive selection, α (note this is different from the α estimated in our linked selection models above), of 0-fold non-synonymous mutations using the Asymptotic extension of the McDonald–Kreitman (MK) test (Messer and Petrov 2013), where α is calculated at each allele frequency bin of the uSFS (from $1/n$ to $(n-1)/n$). At each allele frequency bin, each α was calculated as

$$\alpha = 1 - \frac{d_0}{d}\frac{p}{p_0}$$

where d_0 and d are the number of derived fixed differences for selected and putatively neutral sites, respsectively, and p_0 and p are the number of selected and putatively neutral polymorphic sites. We identified 0-fold and 4-fold sites using Python (https://github.com/silastittes/cds_fold). We fit the Asymptotic MK extension as a nonlinear Bayesian mixed-model using the R package *brms* (Bürkner 2017, 2018).

$$\alpha_{ij} \sim a_j + b_j e^{-c_j x_{ij}}$$

where α_{ij} is the value of α calculated at the ith allele frequency bin of the jth population and x_{ij} is the corresponding allele frequency bin. The nonlinear brms model was coded as

```
alpha ~
    a + b * exp(-c * allele_frequency)
```

where all three free parameters of the asymptotic function (a, b, and c) were treated as random effects of population and nucleotide type (see Supplement IV), and the subspecies was treated as a fixed effect. These effects were coded in brms as

```
a + b + c ~
    1 + allele_frequency +
    (1 + allele_frequency | pop) +
    (1 + allele_frequency | nuc_type) +
    ssp
```

Identifying Selective Sweeps

We used RAiSD (Alachiotis and Pavlidis 2018) to infer regions that have experienced a selective sweep in each population, including the rangewide samples. Across all populations, we used a minor allele frequency threshold of 0.05 and a window size of 100 SNPs. We called SNPs and generated VCF files for each population using the *dovcf* utility from ANGSD, using a SNP calling p value cut off of 1×10^{-6}. We used the Python package mop (https://pypi.org/project/mop-bam/) to exclude SNPs that fell in regions with low and excessive coverage and/or poor quality. Here we required each locus to have at least 70% of individuals with a depth between 5 and 100, and to have a phred scaled quality scores above 30 for base and mapping quality. We additionally used mop to rescale the μ_{var} sub-statistic in each population based on the proportion of high quality bases available in each of the RAiSD SNP windows, after which we recalculated the overall μ statistic as the product of the three sub-statistics (Alachiotis and Pavlidis 2018). After correcting the μ statistic, we defined outliers using simulations of demography using msprime (Kelleher *et al.* 2016). We considered estimates of μ that were greater than the 99.99% quantile of the neutral simulations as evidence for being within a sweep region. RAiSD outliers within 100kb were merged using *Bedtools merge* (Quinlan and Hall 2010) and treated as a single sweep region in downstream analyses. We then used *Bedtools intersect* to identify overlapping sweep regions among populations which we treated as evidence for being a shared sweep.

Assessing the number of false positive sweeps and the number of shared sweeps expected by chance

To assess false positives, we used a second random sample non-overlapping individuals from the two Palmar Chico populations. False positives were assessed based on the number of sweep regions that did not overlap between the 2 replicate Palmar Chico samples from each subspecies. To account for differences in the total number of sweep regions for each replicate, we averaged the two proportions

$$P = (\frac{n_S}{n_{P1}} + \frac{n_S}{n_{P2}})/2$$

and

$$FP = 1 - P$$

where n_S is the number of sweeps shared between the replicates, and n_{P1} and n_{P2} were the number of sweeps in the first and second replicates, respectively. In downstream analyses, we used the average value of FP for the two subspecies, although it was higher for maize than for teosinte (0.60 and 0.49 for maize and teosinte, respectively).

We evaluated the number of sweeps that we would expect populations to share by chance using a simple statistical test based on the hypergeometric distribution,

$$Pr(x \geq X) = \sum_{x \geq X}^{n_1} \frac{\binom{n_2}{x}\binom{N}{n_1 - x}}{\binom{N}{n_1}}$$

where N is the total number of loci tested; n_1 and n_2 are the number of outlier loci in the first and second populations, respectively; and x is the number of shared outliers between the two populations. The population with the larger number of outliers was always designated at the first population. We accounted for false positive by multiplying the raw values of N, n_1, n_2, and x by the FP value described above.

We corrected p-values for multiple tests using the Benjamini and Yekutieli method, implemented with the R function, *p.adjust* (Benjamini and Yekutieli 2001).

Inferring modes of convergent adaptation

For sweep regions that were overlapping in 2 to 9 of the 11 populations, we used *rdmc* to infer the most likely mode of convergent adaptation (Lee and Coop 2017; Tittes 2020). To ensure a sufficient number loci were included to estimate decay in covariance across the sweep regions, we added 10% of each sweep region's total length on each of its ends prior to fitting the models. To reduce the computation time, we exclude sites that had an allele frequency less than 1/20 across all populations. As sweep regions differed in size, we subset from their total number of sites to maintain approximately 250K SNPs per centiMorgan. All sweeps regions were required to have fewer than 100K SNPs. Sweep regions near the ends of chromosomes for which we could not estimate the number of centiMorgans for were subset to 10K SNPs. To contrast allele frequency covariance in sweep regions, we used allele frequencies at 100K random loci to estimate the neutral covariance among all populations. To reduce computation time when fitting the models, we estimated a covariance matrix using a random subset of 50K of these loci. When fitting *rdmc*, we assumed the effective population size was 50K. The recombination rate was approximated for each sweep region as the median interpolated value based on a previously generated genetic map for maize that was lifted on to the v5 reference genome (Ogut *et al.* 2015). The *rdmc* function arguments that control the grid of parameter values over which composite likelihood were computed were:

```
n_sites = 20,
num_bins = 1000,
sels = 10^seq(-5, -1, length.out = 15),
times = c(1e2, 1e3, 1e4, 1e5),
gs = 10^seq(-3, -1, length.out = 3),
migs = 10^(seq(-3, -1, length.out = 2)),
cholesky = TRUE
```

We compared composite likelihoods over four convergent adaptation modes, "neutral", "independent", "standing", and "migration". We assigned each sweep to the mode with the highest log composite likelihood. To assess the overall performance of the method to distinguish between the four modes, we computed differences between the highest composite likelihood and the next highest for each sweep.

All wrangling to prepare input data for statistical analyses was done using *R* (R Core Team 2020) with appreciable reliance on functions from the tidyverse package suite (Wickham *et al.* 2019). Figures were made using ggplot2 (Wickham 2016), patchwork (Pedersen 2019), and cowplot (Wilke 2019). Code and summarised data for the entirety of the analyses, including Jupyter notebooks for reproducing figures, is available from https://github.com/silastittes/parv_local permanent repository to follow.

Acknowledgments

We would like to thank Andi Kur for providing the corn art, along with Matthew Gibson, Tom Booker, Cathy Rushworth, and other members of the Ross-Ibarra lab for feedback and suggestions on early drafts of the manuscript. This work was funded in part by grants from the National Science Foundation (1822330 and 1238014). We would also like to acknowledge Felix Andrews for statistical advice, although we did not follow it.

References

Aguirre-Liguori, J. A., B. S. Gaut, J. P. Jaramillo-Correa, M. I. Tenaillon, S. Montes-Hernández, *et al.*, 2019 Divergence with gene flow is driven by local adaptation to temperature and soil phosphorus concentration in teosinte subspecies (zea mays parviglumis and zea mays mexicana). Molecular ecology **28**: 2814–2830.

Alachiotis, N. and P. Pavlidis, 2018 Raisd detects positive selection based on multiple signatures of a selective sweep and snp vectors. Communications biology **1**: 1–11.

Beissinger, T. M., L. Wang, K. Crosby, A. Durvasula, M. B. Hufford, *et al.*, 2016 Recent demography drives changes in linked selection across the maize genome. Nature plants **2**: 1–7.

Bellon, M. R., A. Mastretta-Yanes, A. Ponce-Mendoza, D. Ortiz-Santamaría, O. Oliveros-Galindo, *et al.*, 2018 Evolutionary and food supply implications of ongoing maize domestication by mexican campesinos. Proceedings of the Royal Society B **285**: 20181049.

Benjamini, Y. and D. Yekutieli, 2001 The control of the false discovery rate in multiple testing under dependency. Annals of statistics pp. 1165–1188.

Bourne, E. C., G. Bocedi, J. M. Travis, R. J. Pakeman, R. W. Brooker, *et al.*, 2014 Between migration load and evolutionary rescue: dispersal, adaptation and the response of spatially structured populations to environmental change. Proceedings of the Royal Society B: Biological Sciences **281**: 20132795.

Bragg, J. G., M. A. Supple, R. L. Andrew, and J. O. Borevitz, 2015 Genomic variation across landscapes: insights and applications. New Phytologist **207**: 953–967.

Browning, B. L. and S. R. Browning, 2013 Detecting identity by descent and estimating genotype error rates in sequence data. The American Journal of Human Genetics **93**: 840–851.

Buckler, E. and T. P. Holtsford, 1996 Zea systematics: ribosomal its evidence. Molecular Biology and Evolution **13**: 612–622.

Buffalo, V., 2021 Why do species get a thin slice of π? revisiting lewontin's paradox of variation. bioRxiv .

Butler, E. E. and P. Huybers, 2013 Adaptation of us maize to temperature variations. Nature Climate Change **3**: 68–72.

Bürkner, P.-C., 2017 brms: An R package for Bayesian multilevel models using Stan. Journal of Statistical Software **80**: 1–28.

Bürkner, P.-C., 2018 Advanced Bayesian multilevel modeling with the R package brms. The R Journal **10**: 395–411.

Calfee, E., D. Gates, A. Lorant, M. T. Perkins, G. Coop, *et al.*, 2021 Selective sorting of ancestral introgression in maize and teosinte along an elevational cline. bioRxiv .

Charlesworth, B., 2020 How long does it take to fix a favorable mutation, and why should we care? The American Naturalist **195**: 753–771.

Chen, Q., L. F. Samayoa, C. J. Yang, P. J. Bradbury, B. A. Olukolu, *et al.*, 2020 The genetic architecture of the maize progenitor, teosinte, and how it was altered during maize domestication. PLoS genetics **16**: e1008791.

Chevin, L.-M., G. Martin, and T. Lenormand, 2010 Fisher's model and the genomics of adaptation: restricted pleiotropy, heterogenous mutation, and parallel evolution. Evolution: International Journal of Organic Evolution **64**: 3213–3231.

Clark, R. M., S. Tavaré, and J. Doebley, 2005 Estimating a nucleotide substitution rate for maize from polymorphism at a major domestication locus. Molecular biology and evolution **22**: 2304–2312.

Clausen, J., D. D. Keck, W. M. Hiesey, *et al.*, 1948 Experimental studies on the nature of species. iii. environresponses of climatic races of achillea. Experimental studies on the nature of species. III. Environresponses of climatic races of Achillea. .

Colosimo, P. F., K. E. Hosemann, S. Balabhadra, G. Villarreal, M. Dickson, *et al.*, 2005 Widespread parallel evolution in sticklebacks by repeated fixation of ectodysplasin alleles. science **307**: 1928–1933.

Coop, G. and P. Ralph, 2012 Patterns of neutral diversity under general models of selective sweeps. Genetics **192**: 205–224.

Corbett-Detig, R. B., D. L. Hartl, and T. B. Sackton, 2015 Natural selection constrains neutral diversity across a wide range of species. PLoS Biol **13**: e1002112.

Coventry, A., L. M. Bull-Otterson, X. Liu, A. G. Clark, T. J. Maxwell, *et al.*, 2010 Deep resequencing reveals excess rare recent variants consistent with explosive population growth. Nature communications **1**: 1–6.

DeWitt, W. S., K. D. Harris, A. P. Ragsdale, and K. Harris, 2021 Nonparametric coalescent inference of mutation spectrum history and demography. Proceedings of the National Academy of Sciences **118**.

Doebley, J., 1989 Isozymic evidence and the evolution of crop plants. In *Isozymes in plant biology*, pp. 165–191, Springer.

Ellstrand, N. C., L. C. Garner, S. Hegde, R. Guadagnuolo, and L. Blancas, 2007 Spontaneous hybridization between maize and teosinte. Journal of Heredity **98**: 183–187.

Eyre-Walker, A., R. L. Gaut, H. Hilton, D. L. Feldman, and B. S. Gaut, 1998 Investigation of the bottleneck leading to the domestication of maize. Proceedings of the National Academy of Sciences **95**: 4441–4446.

Fournier-Level, A., A. Korte, M. D. Cooper, M. Nordborg, J. Schmitt, *et al.*, 2011 A map of local adaptation in arabidopsis thaliana. Science **334**: 86–89.

Freed, D., R. Aldana, J. A. Weber, and J. S. Edwards, 2017 The sentieon genomics tools-a fast and accurate solution to variant calling from next-generation sequence data. BioRxiv p. 115717.

Fustier, M.-A., J.-T. Brandenburg, S. Boitard, J. Lapeyronnie, L. Eguiarte, *et al.*, 2017 Signatures of local adaptation in lowland and highland teosintes from whole-genome sequencing of pooled samples. Molecular Ecology **26**: 2738–2756.

Fustier, M.-A., N. E. Martínez-Ainsworth, J. A. Aguirre-Liguori, A. Venon, H. Corti, *et al.*, 2019 Common gardens in teosintes reveal the establishment of a syndrome of adaptation to altitude. PLoS genetics **15**: e1008512.

Galtier, N., 2016 Adaptive protein evolution in animals and the effective population size hypothesis. PLoS genetics **12**: e1005774.

Gates, D. J., D. Runcie, G. M. Janzen, A. R. Navarro, M. Willcox, *et al.*, 2019 Single-gene resolution of locally adaptive genetic variation in mexican maize. BioRxiv p. 706739.

Gillespie, J. H., 2004 *Population genetics: a concise guide*. JHU Press.

Gossmann, T. I., B.-H. Song, A. J. Windsor, T. Mitchell-Olds, C. J. Dixon, *et al.*, 2010 Genome wide analyses reveal little evidence for adaptive evolution in many plant species. Molecular biology and evolution **27**: 1822–1832.

Haines, M. R., M. R. Haines, and R. H. Steckel, 2000 *A population history of North America*. Cambridge University Press.

Haller, B. C. and P. W. Messer, 2019 Slim 3: forward genetic simulations beyond the wright–fisher model. Molecular biology and evolution **36**: 632–637.

Hämälä, T. and P. Tiffin, 2020 Biased gene conversion constrains adaptation in arabidopsis thaliana. Genetics **215**: 831–846.

Hamrick, J. and R. W. Allard, 1972 Microgeographical variation in allozyme frequencies in avena barbata. Proceedings of the National Academy of Sciences **69**: 2100–2104.

Hanghøj, K., I. Moltke, P. A. Andersen, A. Manica, and T. S. Korneliussen, 2019 Fast and accurate relatedness estimation from high-throughput sequencing data in the presence of inbreeding. Gigascience **8**: giz034.

Hoban, S., J. L. Kelley, K. E. Lotterhos, M. F. Antolin, G. Bradburd, *et al.*, 2016 Finding the genomic basis of local adaptation: pitfalls, practical solutions, and future directions. The American Naturalist **188**: 379–397.

Hudson, R. R. and N. L. Kaplan, 1995 Deleterious background selection with recombination. Genetics **141**: 1605–1617.

Hufford, M. B., 2010 *Genetic and ecological approaches to guide conservation of teosinte (Zea mays ssp. parviglumis), the wild progenitor of maize*. University of California, Davis.

Hufford, M. B., A. S. Seetharam, M. R. Woodhouse, K. M. Chougule, S. Ou, *et al.*, 2021 De novo assembly, annotation, and comparative analysis of 26 diverse maize genomes. Science .

Hufford, M. B., X. Xu, J. Van Heerwaarden, T. Pyhäjärvi, J.-M. Chia, *et al.*, 2012 Comparative population genomics of maize domestication and improvement. Nature genetics **44**: 808.

Keinan, A. and A. G. Clark, 2012 Recent explosive human population growth has resulted in an excess of rare genetic variants. science **336**: 740–743.

Kelleher, J., A. M. Etheridge, and G. McVean, 2016 Efficient coalescent simulation and genealogical analysis for large sample sizes. PLoS computational biology **12**: e1004842.

Korneliussen, T. S., A. Albrechtsen, and R. Nielsen, 2014 Angsd: analysis of next generation sequencing data. BMC bioinformatics **15**: 356.

Lee, K. M. and G. Coop, 2017 Distinguishing among modes of convergent adaptation using population genomic data. Genetics **207**: 1591–1619.

Levy, R. and E. Borenstein, 2012 Reverse ecology: from systems to environments and back. In *Evolutionary systems biology*, pp. 329–345, Springer.

Lewontin, R. C. *et al.*, 1974 *The genetic basis of evolutionary change*, volume 560. Columbia University Press New York.

Li, H., 2013 Aligning sequence reads, clone sequences and assembly

contigs with bwa-mem. arXiv preprint arXiv:1303.3997 .
Li, Y. F., J. C. Costello, A. K. Holloway, and M. W. Hahn, 2008 "reverse ecology" and the power of population genomics. Evolution: International Journal of Organic Evolution **62**: 2984–2994.
Lowry, D. B., R. C. Rockwood, and J. H. Willis, 2008 Ecological reproductive isolation of coast and inland races of mimulus guttatus. Evolution: International Journal of Organic Evolution **62**: 2196–2214.
Mei, W., M. G. Stetter, D. J. Gates, M. C. Stitzer, and J. Ross-Ibarra, 2018 Adaptation in plant genomes: Bigger is different.
Messer, P. W. and D. A. Petrov, 2013 Frequent adaptation and the mcdonald–kreitman test. Proceedings of the National Academy of Sciences **110**: 8615–8620.
Nannas, N. J. and R. K. Dawe, 2015 Genetic and genomic toolbox of zea mays. Genetics **199**: 655–669.
O'Brien, A. M., R. J. Sawers, S. Y. Strauss, and J. Ross-Ibarra, 2019 Adaptive phenotypic divergence in an annual grass differs across biotic contexts. Evolution **73**: 2230–2246.
Ogut, F., Y. Bian, P. J. Bradbury, and J. B. Holland, 2015 Joint-multiple family linkage analysis predicts within-family variation better than single-family analysis of the maize nested association mapping population. Heredity **114**: 552–563.
Orozco-Ramírez, Q., H. Perales, and R. J. Hijmans, 2017 Geographical distribution and diversity of maize (zea mays l. subsp. mays) races in mexico. Genetic resources and crop evolution **64**: 855–865.
Ossowski, S., K. Schneeberger, J. I. Lucas-Lledó, N. Warthmann, R. M. Clark, *et al.*, 2010 The rate and molecular spectrum of spontaneous mutations in arabidopsis thaliana. science **327**: 92–94.
Pedersen, T. L., 2019 *patchwork: The Composer of Plots*. R package version 1.0.0.
Pickrell, J. and J. Pritchard, 2012 Inference of population splits and mixtures from genome-wide allele frequency data. Nature Precedings pp. 1–1.
Piperno, D. R., 1991 The status of phytolith analysis in the american tropics. Journal of World Prehistory **5**: 155–191.
Piperno, D. R., A. J. Ranere, I. Holst, J. Iriarte, and R. Dickau, 2009 Starch grain and phytolith evidence for early ninth millennium bp maize from the central balsas river valley, mexico. Proceedings of the National Academy of Sciences **106**: 5019–5024.
Portwood, J. L., M. R. Woodhouse, E. K. Cannon, J. M. Gardiner, L. C. Harper, *et al.*, 2019 Maizegdb 2018: the maize multi-genome genetics and genomics database. Nucleic acids research **47**: D1146–D1154.
Pritchard, J. K., J. K. Pickrell, and G. Coop, 2010 The genetics of human adaptation: hard sweeps, soft sweeps, and polygenic adaptation. Current biology **20**: R208–R215.
Pyhäjärvi, T., M. B. Hufford, S. Mezmouk, and J. Ross-Ibarra, 2013 Complex patterns of local adaptation in teosinte. Genome biology and evolution **5**: 1594–1609.
Quinlan, A. R. and I. M. Hall, 2010 Bedtools: a flexible suite of utilities for comparing genomic features. Bioinformatics **26**: 841–842.
R Core Team, 2020 *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
Ranum, P., J. P. Peña-Rosas, and M. N. Garcia-Casal, 2014 Global maize production, utilization, and consumption. Annals of the new York academy of sciences **1312**: 105–112.
Rodriguez-Zapata, F., A. C. Barnes, K. A. Blocher-Juarez, D. J. Gates, A. Kur, *et al.*, 2021 Teosinte introgression modulates phosphatidylcholine levels and induces early maize flowering time. bioRxiv .
Ross-Ibarra, J., M. Tenaillon, and B. S. Gaut, 2009 Historical divergence and gene flow in the genus zea. Genetics **181**: 1399–1413.
Rudman, S. M., S. I. Greenblum, S. Rajpurohit, N. J. Betancourt, J. Hanna, *et al.*, 2021 Direct observation of adaptive tracking on ecological timescales in drosophila. bioRxiv .
Savolainen, O., M. Lascoux, and J. Merilä, 2013 Ecological genomics of local adaptation. Nature Reviews Genetics **14**: 807–820.
Schrider, D. R. and A. D. Kern, 2016 S/hic: robust identification of soft and hard sweeps using machine learning. PLoS genetics **12**: e1005928.
Skotte, L., T. S. Korneliussen, and A. Albrechtsen, 2013 Estimating individual admixture proportions from next generation sequencing data. Genetics **195**: 693–702.
Smith, N. G. and A. Eyre-Walker, 2002 Adaptive protein evolution in drosophila. Nature **415**: 1022–1024.
Stitzer, M. C. and J. Ross-Ibarra, 2018 Maize domestication and gene interaction. New phytologist **220**: 395–408.
Swarts, K., R. M. Gutaker, B. Benz, M. Blake, R. Bukowski, *et al.*, 2017 Genomic estimation of complex traits reveals ancient maize adaptation to temperate north america. Science **357**: 512–515.
Tenaillon, M. I., M. B. Hufford, B. S. Gaut, and J. Ross-Ibarra, 2011 Genome size and transposable element content as determined by high-throughput sequencing in maize and zea luxurians. Genome biology and evolution **3**: 219–229.
Tittes, S., 2020 rdmc: an open source r package implementing convergent adaptation models of lee and coop (2017). G3: Genes, Genomes, Genetics **10**: 3041–3046.
Tuanmu, M.-N. and W. Jetz, 2015 A global, remote sensing-based characterization of terrestrial habitat heterogeneity for biodiversity and ecosystem modelling. Global Ecology and Biogeography **24**: 1329–1339.
Uricchio, L. H., D. A. Petrov, and D. Enard, 2019 Exploiting selection at linked sites to infer the rate and strength of adaptation. Nature ecology & evolution **3**: 977–984.
Van Heerwaarden, J., J. ROSS-IBARRA, J. Doebley, J. C. Glaubitz, J. DE JESÚS SÁNCHEZ GONZÁLEZ, *et al.*, 2010 Fine scale genetic structure in the wild ancestor of maize (zea mays ssp. parviglumis). Molecular Ecology **19**: 1162–1173.
Wagenmakers, E.-J. and S. Farrell, 2004 Aic model selection using akaike weights. Psychonomic bulletin & review **11**: 192–196.
Wang, L., T. M. Beissinger, A. Lorant, C. Ross-Ibarra, J. Ross-Ibarra, *et al.*, 2017 The interplay of demography and selection during maize domestication and expansion. Genome biology **18**: 1–13.
Wang, L., E. B. Josephs, K. M. Lee, L. M. Roberts, R. Rellán-Álvarez, *et al.*, 2020 Molecular parallelism underlies convergent highland adaptation of maize landraces. bioRxiv .
Whitehead, A., F. Galvez, S. Zhang, L. M. Williams, and M. F. Oleksiak, 2011 Functional genomics of physiological plasticity and local adaptation in killifish. Journal of Heredity **102**: 499–511.
Wickham, H., 2016 *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.
Wickham, H., M. Averick, J. Bryan, W. Chang, L. D. McGowan, *et al.*, 2019 Welcome to the tidyverse. Journal of Open Source Software **4**: 1686.
Wilke, C. O., 2019 *cowplot: Streamlined Plot Theme and Plot Annotations for 'ggplot2'*. R package version 1.0.0.
Wilkes, H. G. *et al.*, 1967 Teosinte: The closest relative of maize. Teosinte: the closest relative of maize. .
Wright, S. I., I. V. Bi, S. G. Schroeder, M. Yamasaki, J. F. Doebley, *et al.*, 2005 The effects of artificial selection on the maize genome. Science **308**: 1310–1314.
Xue, A. T., D. R. Schrider, and A. D. Kern, 2021 Discovery of ongoing selective sweeps within anopheles mosquito populations using deep learning. Molecular biology and evolution **38**: 1168–1183.

Supplement I

Population sampling locations

Table S1 Population sampling location information.

Population	Subspecies	Sample size	Latitude	Longitude	Elevation (meters)	Accession ID
Crucero Lagunitas	Maize	10	16.98	-99.28	201	2373-GRO-294
Amatlán de Cañas	Maize	10	20.82	-104.41	760	5054-NAY-310
Los Guajes	Maize	10	19.23	-100.49	985	TC-300
San Lorenzo	Maize	10	19.94	-103.99	982	RMM-15
Palmar Chico	Maize	55	18.64	-100.35	1008	JSG-RMM-LCL-529
Crucero Lagunitas	Teosinte	10	16.85	-99.06	590	JSG-RMM-LCL-487
Amatlán de Cañas	Teosinte	10	20.82	-104.41	880	JSG-JRP-ERG-543
El Rodeo	Teosinte	10	16.35	-97.02	982	JSG-RMM-LCL-486
Los Guajes	Teosinte	10	19.23	-100.49	851	JSG Y RMM-454
San Lorenzo	Teosinte	10	19.94	-103.99	982	RMM-13
Palmar Chico	Teosinte	50	18.64	-100.35	983	JSG-RMM-LCL-528

Supplement II

Further assessment of f4 statistic inferences

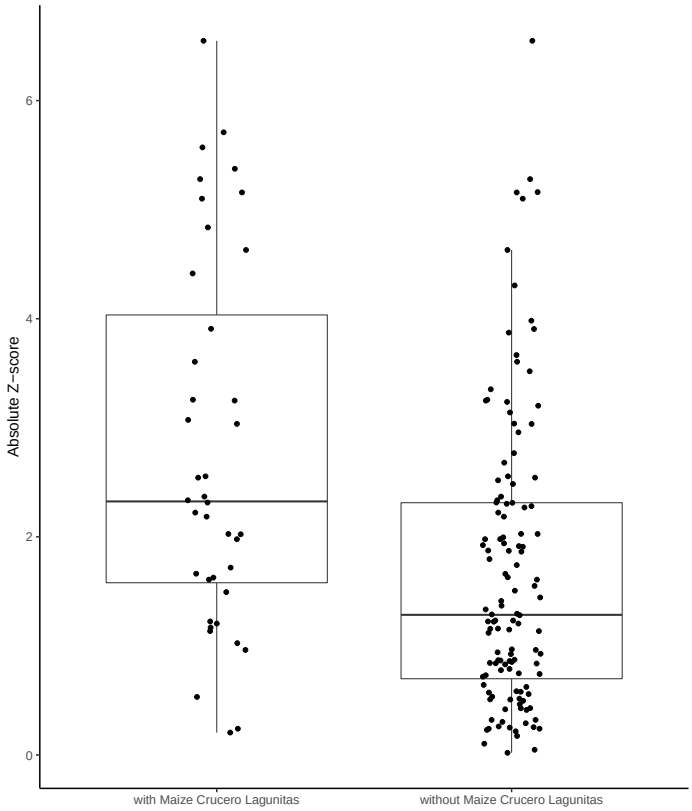

Figure S1 **f4 tests including the maize Crucero Lagunitas population are significantly elevated compared to those without.**

Table S2 **Significant F4 tests. Each row of the table reports the number of significant f4 tests that occurred with a given focal and secondary population, where the two other tip positions were filled with each of the remaining populations for each subspecies. Rows that are left blank in the secondary column are used to report the total number of significant trees for a given focal population.**

Focal population	Secondary population	Count
Maize Amatlan de Canas		5
Maize Amatlan de Canas	Maize Crucero Lagunitas	5
Maize Amatlan de Canas	Teosinte Amatlan de Canas	3
Maize Amatlan de Canas	Teosinte El Rodeo	2
Maize Amatlan de Canas	Teosinte Palmar Chico	2
Maize Amatlan de Canas	Teosinte San Lorenzo	2
Maize Amatlan de Canas	Teosinte Los Guajes	1
Maize Crucero Lagunitas		15
Maize Crucero Lagunitas	Teosinte Amatlan de Canas	9
Maize Crucero Lagunitas	Teosinte Crucero Lagunitas	6
Maize Crucero Lagunitas	Teosinte El Rodeo	6
Maize Crucero Lagunitas	Maize Palmar Chico	5
Maize Crucero Lagunitas	Maize Los Guajes	4
Maize Crucero Lagunitas	Teosinte San Lorenzo	4

Maize Crucero Lagunitas	Maize Amatlan de Canas	3
Maize Crucero Lagunitas	Maize San Lorenzo	3
Maize Crucero Lagunitas	Teosinte Palmar Chico	3
Maize Crucero Lagunitas	Teosinte Los Guajes	2
Maize Los Guajes		6
Maize Los Guajes	Teosinte Amatlan de Canas	4
Maize Los Guajes	Maize Crucero Lagunitas	3
Maize Los Guajes	Teosinte Crucero Lagunitas	3
Maize Los Guajes	Maize San Lorenzo	2
Maize Los Guajes	Teosinte Palmar Chico	2
Maize Los Guajes	Maize Palmar Chico	1
Maize Los Guajes	Teosinte El Rodeo	1
Maize Los Guajes	Teosinte Los Guajes	1
Maize Los Guajes	Teosinte San Lorenzo	1
Maize Palmar Chico		9
Maize Palmar Chico	Teosinte Amatlan de Canas	7
Maize Palmar Chico	Maize Crucero Lagunitas	5
Maize Palmar Chico	Teosinte Palmar Chico	4
Maize Palmar Chico	Teosinte El Rodeo	3
Maize Palmar Chico	Maize Los Guajes	2
Maize Palmar Chico	Maize San Lorenzo	2
Maize Palmar Chico	Teosinte San Lorenzo	2
Maize Palmar Chico	Teosinte Crucero Lagunitas	1
Maize Palmar Chico	Teosinte Los Guajes	1
Maize San Lorenzo		6
Maize San Lorenzo	Teosinte Amatlan de Canas	4
Maize San Lorenzo	Maize Crucero Lagunitas	3
Maize San Lorenzo	Maize Los Guajes	2
Maize San Lorenzo	Teosinte Crucero Lagunitas	2
Maize San Lorenzo	Teosinte El Rodeo	2
Maize San Lorenzo	Teosinte Palmar Chico	2
Maize San Lorenzo	Maize Palmar Chico	1
Maize San Lorenzo	Teosinte Los Guajes	1
Maize San Lorenzo	Teosinte San Lorenzo	1
Teosinte Amatlan de Canas		11
Teosinte Amatlan de Canas	Maize Crucero Lagunitas	9
Teosinte Amatlan de Canas	Maize Palmar Chico	4
Teosinte Amatlan de Canas	Maize Amatlan de Canas	3
Teosinte Amatlan de Canas	Maize Los Guajes	3
Teosinte Amatlan de Canas	Maize San Lorenzo	3
Teosinte Amatlan de Canas	Teosinte Los Guajes	3
Teosinte Amatlan de Canas	Teosinte Palmar Chico	3
Teosinte Amatlan de Canas	Teosinte San Lorenzo	3
Teosinte Amatlan de Canas	Teosinte Crucero Lagunitas	1
Teosinte Amatlan de Canas	Teosinte El Rodeo	1
Teosinte Crucero Lagunitas		9
Teosinte Crucero Lagunitas	Maize Crucero Lagunitas	6
Teosinte Crucero Lagunitas	Maize Los Guajes	5
Teosinte Crucero Lagunitas	Teosinte Amatlan de Canas	4
Teosinte Crucero Lagunitas	Teosinte El Rodeo	4
Teosinte Crucero Lagunitas	Maize Palmar Chico	3

Teosinte Crucero Lagunitas	Maize San Lorenzo	3
Teosinte Crucero Lagunitas	Maize Amatlan de Canas	1
Teosinte Crucero Lagunitas	Teosinte Palmar Chico	1
Teosinte Los Guajes	Maize Crucero Lagunitas	3
Teosinte Los Guajes	Teosinte Amatlan de Canas	3
Teosinte Los Guajes		3
Teosinte Los Guajes	Maize Los Guajes	1
Teosinte Los Guajes	Maize Palmar Chico	1
Teosinte Los Guajes	Maize San Lorenzo	1
Teosinte Palmar Chico		8
Teosinte Palmar Chico	Maize Crucero Lagunitas	5
Teosinte Palmar Chico	Teosinte Amatlan de Canas	5
Teosinte Palmar Chico	Maize Palmar Chico	4
Teosinte Palmar Chico	Maize Los Guajes	3
Teosinte Palmar Chico	Maize San Lorenzo	3
Teosinte Palmar Chico	Teosinte El Rodeo	2
Teosinte Palmar Chico	Maize Amatlan de Canas	1
Teosinte Palmar Chico	Teosinte Crucero Lagunitas	1
Teosinte San Lorenzo	Maize Crucero Lagunitas	5
Teosinte San Lorenzo		5
Teosinte San Lorenzo	Teosinte Amatlan de Canas	3
Teosinte San Lorenzo	Maize Palmar Chico	2
Teosinte San Lorenzo	Teosinte El Rodeo	2
Teosinte San Lorenzo	Maize Amatlan de Canas	1
Teosinte San Lorenzo	Maize Los Guajes	1
Teosinte San Lorenzo	Maize San Lorenzo	1

Supplement III

Validating empirical relationship between α and θ.

To better understand and validate the observed empirical relationship between α, the proportion of new mutations fixed by natural selection, and θ, the population scaled mutation rate ($4N_e\mu$), we conducted simulations using SLiM (Haller and Messer 2019). Each simulation consisted of a 20Mb sequence with constant recombination and mutation rates of 1×10^6 and 3×10^6, respectively, both of which are 100-fold higher than our estimated rates in maize. Matching the average values for the maize genome, the simulated 20Mb sequence had 300 gene regions separated by 30Kb each. Each gene region consisted of four 200bp exons and three 300bp introns. Mutations were only simulated within exons. Fitness effects of mutations were simulated to be neutral, strictly deleterious, or strictly positive, where the two non-neutral mutation classes were drawn from independent gamma distributions with mean and shape parameters (μ_+, $shape_+$, and μ_-, $shape_-$, for the positive and negative distributions, respectively). Following the empirical ratio of 0-fold nonsynonymous to 4-fold synonymous sites, the proportion of non-neutral mutations to neutral mutations was simulated as 3.5 to 1. The ratio of positively to negatively selected mutations (p_-) was allowed to vary across simulations. We implemented as simple 3-epoch demographic model where the three population sizes ("ancestral" $= N_A$, "bottleneck"$= N_B$, and "modern" $= N_0$) and the generation times ("bottleneck" $= T_B$ and "modern" = T_0) for changing to a new epoch were allowed to vary across simulations. For each simulation, we computed $\alpha = \frac{\rho_+}{\rho_+ + \rho_- + \rho_0}$, where ρ_+, ρ_-, ρ_0, are the frequency of positive, negative, and neutral fixations that occurred over the course of the simulation. Simulations were always initiated with a biurn-in of $10 * N_A$ generations to ensure simulations were near equilibrium conditions prior to implementing demographic changes. To make direct comparisons with the empirical data, we converted the demographic parameters to a single value of θ as the harmonic mean of the three population size parameters, weighted by the number of generations spent at that population size. This resulted in 7 free parameters that were drawn from uniform priors (Table S3). From 1000 simulations, we fit a linear model predicting α from the free parameters.

As expected from theory, we found a positive effect of the population scaled mutation rate ($\theta = 4N_e\mu$) on α (Table S3 and Figure S2), and the total number of adaptive substitutions (Table S4 and Figure S2). However, in both cases, the relationship appears to be nonlinear and plateaus for higher values of θ. The non-linearity persists after accounting for the effect of other parameters (Figure S2 C and D). Nonetheless, α and the number of adaptive substitutions increase over the broad range of θ simulated.

Unsurprisingly, the mean of both positive and negative fitness distributions, and the proportion of non-neutral mutations simulated as deleterious, were also all significant predictors of α. As noted in the discussion, given the relative ease of fixing strongly adaptive mutations, it is now widely believed that difference in the ability to purge weakly deleterious alleles plays an important role in determining α. Our simulations were conducted over a broad distribution of fitness effects. As such, we cannot directly identify to what extent the relationship between α and θ is driven by increasing efficacy of larger populations to fix positive mutations versus their ability to purge weekly deleterious ones. However, based on our simulations, it seems likely that both phenomena contribute to α, though which is more important will depend on the particular combination of parameters that a population occupies.

Table S3 Regression predicting α from demographic and fitness effect parameters. The Prior column contains the distributions that each parameter was simulated from. The remaining parameters are those output from the *lm* function in *R*. For each simulation, θ was calculated as the harmonic mean of the three population size parameters, weighted by the number of generations spent at that population size. Priors for the two population size change parameters T_0 and T_B were drawn from a uniform with bound 0 and N_A, but these values were added on the the previous number of generations, such that the total number of generations was $10N_A + T_B + T_0$

Parameter	Prior	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	N/A	3.779	0.197	19.151	1.13e-58
θ	N/A	16.641	1.288	12.923	2.99e-32
N_B	$U(10, 2000)$	N/A	N/A	N/A	N/A
N_A	$U(10, 2000)$	N/A	N/A	N/A	N/A
N_0	$U(55, 2000)$	N/A	N/A	N/A	N/A
T_B	$U(0, N_A)$	N/A	N/A	N/A	N/A
T_0	$U(0, N_A)$	N/A	N/A	N/A	N/A
p_-	$U(0.9, 1)$	-3.999	0.207	-19.314	1.83e-59
μ_+	$U(0, 0.05)$	5.633	0.454	12.394	3.75e-30
$shape_+$	$U(0, 1)$	0.063	0.020	3.147	0.00177
μ_-	$U(-0.05, 0)$	-0.984	0.386	-2.546	0.0113
$shape_-$	$U(0, 1)$	0.125	0.019	6.457	3.04e-10

Table S4 **Regression predicting the count of substitutions $\alpha_{substitutions}$ (as opposed to the proportion of adaptive substitutions in Table S3) from demographic and fitness effect parameters.**

Parameter	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	13.311	0.918	14.493	1.13e-38
θ	120.936	6.045	20.006	1.69e-62
p_-	-11.817	0.972	-12.159	3.14e-29
μ_+	17.709	2.134	8.300	1.55e-15
$shape_+$	0.246	0.094	2.613	0.0093
μ_-	4.686	1.814	2.583	0.0102
$shape_-$	-0.023	0.091	-0.259	0.796

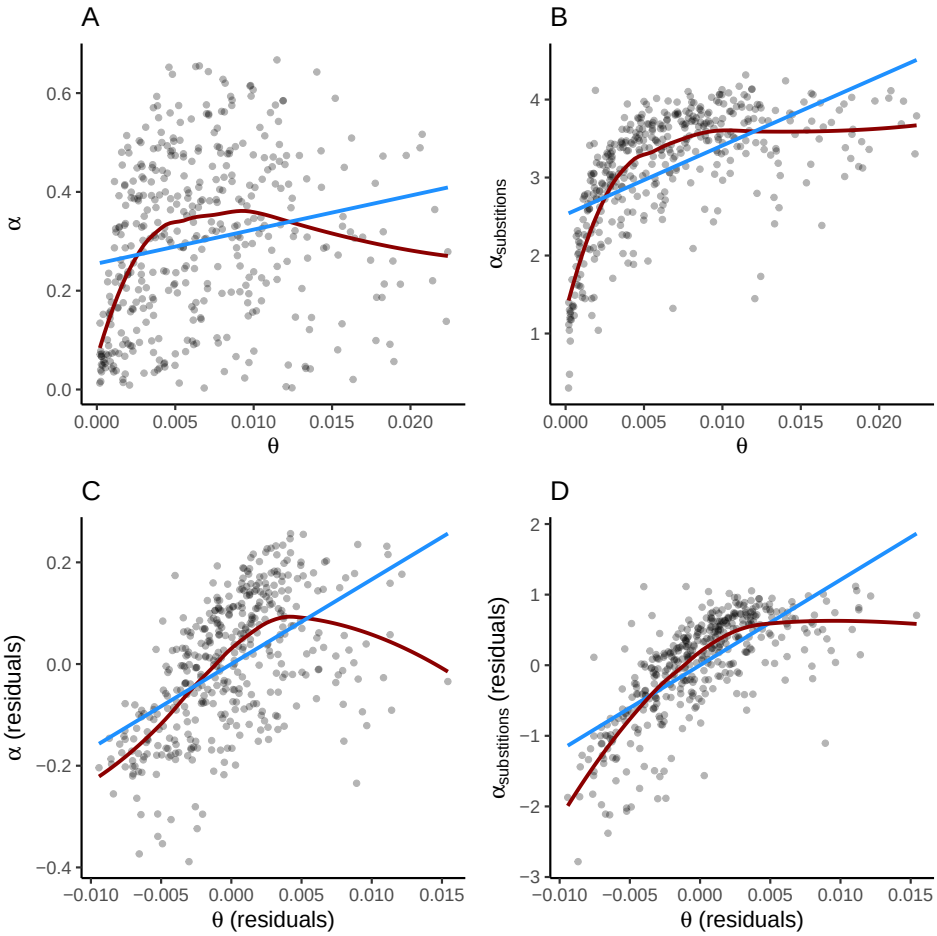

Figure S2 **population scaled mutation rate ($\theta = 4N_e\mu$) is associated with α over a broad range of demography and fitness parameters.** The raw relationship between θ and α (A) and θ and the count of adaptive substitions ($\alpha_{substitutions}$) (B) from simulations, not taking account any associations between α and the other parameters or θ and the other parameters. Added variable (or partial regression) plots showing the relationship between model residuals of θ and α (C) and and θ and $\alpha_{substitutions}$ (D), which take account of the associations between both parameters and the other model predictors (see table S3.) Blue lines for each panel show the linear model fit between the x and y variables, while the dark red lines show the loess smooth fit.

Supplement IV

Predicting α by mutation type

Estimates of α may be effected by differences in the mutation rates of different nucleotides and genomic regions. GC biased gene conversion has been shown to reduce α by making it harder to purge slightly deleterious alleles Hämälä and Tiffin (2020). Likewise, the higher mutation rates observed at methylated cytosine bases increases the rate of $C \rightarrow T$ mutations Ossowski *et al.* (2010), which is another mechanism that could result in variation in α by changing the ability to purging deleterious alleles, or by changing the probability of fixation of new adaptive mutations.

To study this, we used the same approach as Hämälä and Tiffin (2020), where we separated the site frequency spectra based on mutation types according to whether the ancestral and derived nucleotides had a single (weak) or double (strong) hydrogen bond between the DNA strands. As such, we studied three mutations types: $A/T \rightarrow G/C$ mutations (WS), $G/C \rightarrow A/T$ (SW) and $C/G \rightarrow G/C$ or $A/T \rightarrow T/A$ (SS_WW).

Unlike patterns found in *Arabidopsis* (Hämälä and Tiffin 2020), α was highest for WS mutations, although there was considerable overlap between the credible intervals for all mutation types.

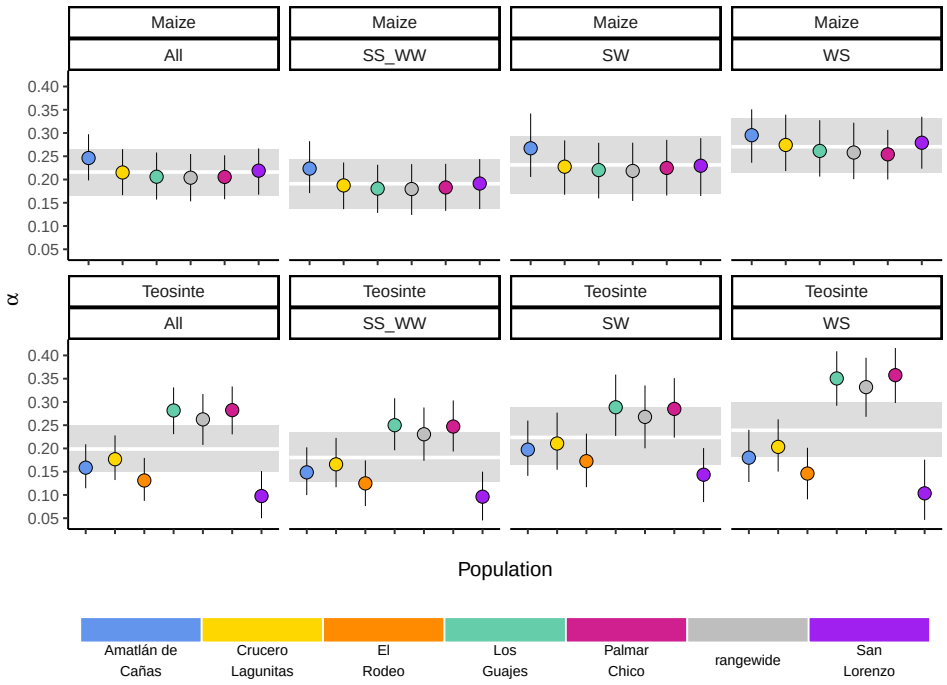

Figure S3 Predicted values of α across mutation types. Grey bands for each mutation type show the 95% credible intervals averaged over each population.

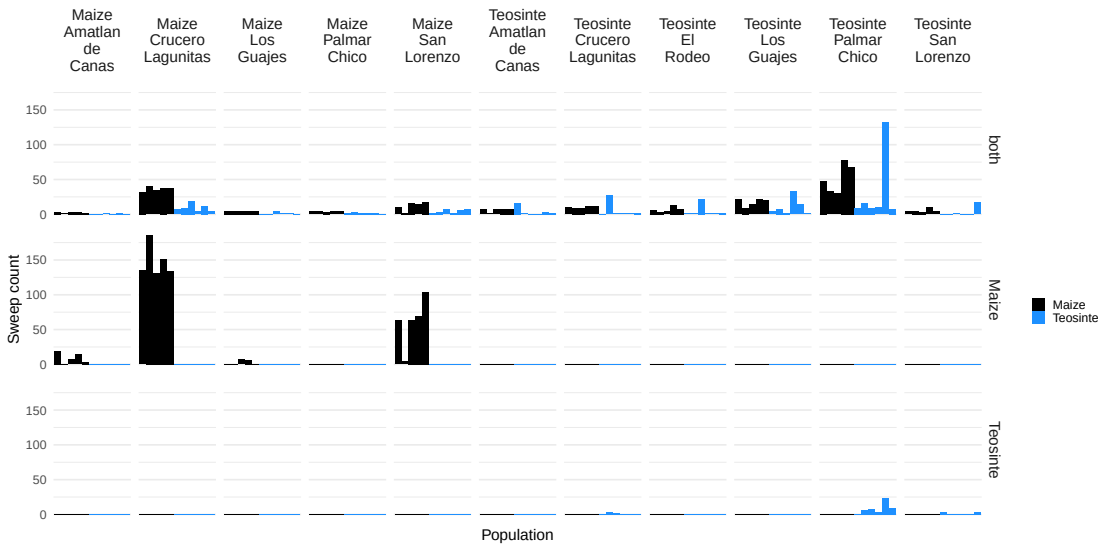

Figure S4 Frequency of each population as the mutation source for sweeps shared via migration. The order of populations along the x axis matches that of the source populations labeled for each strip along the top.

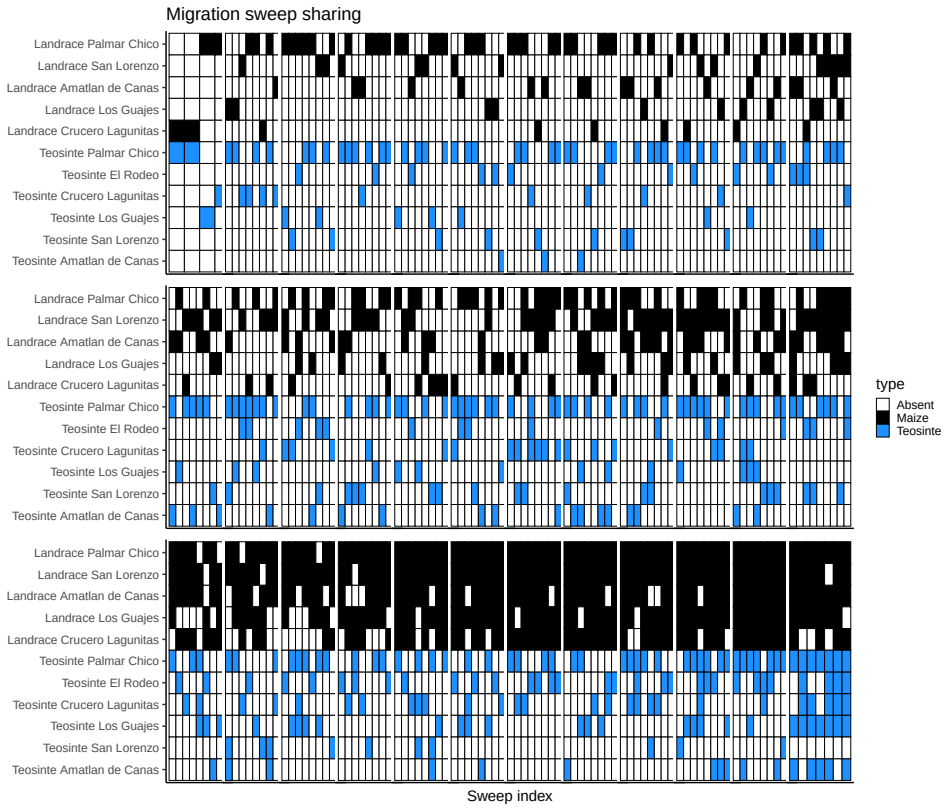

Figure S5 Complete accounting of inferred sweeps shared via migration. The x axis is sorted by the number of populations each sweep was found in. Populations are sorted along the y axis first by subspecies then by their number of sweeps.

Demystifying the high proportion of sweeps unique to the rangewide maize sample

Why does the rangewide sample for maize have such a high proportion of private sweeps? One plausible explanation is that rangewide sweeps represent older sweeps that are shared among all maize, but due to drift and divergence in the individual populations, the sweep signal has been attenuated. When maize individuals are combined from multiple pops, the sweep signal is amplified above the background noise that is generated independently in each population. If this is the case, regions where a sweep unique to the rangewide sample appears should show elevated (but perhaps not statistically significant) evidence of a sweep in the individual populations.

We confirmed this hypothesis by comparing the distribution of the RAiSD μ statistic within the regions we identified as private sweeps in the rangewide samples to a random sample of μ values found genomewide in all the maize populations. Consistent with our expectations, all maize populations have elevated μ values within the regions identified as private sweeps in the rangewide sample. (Figure S6).

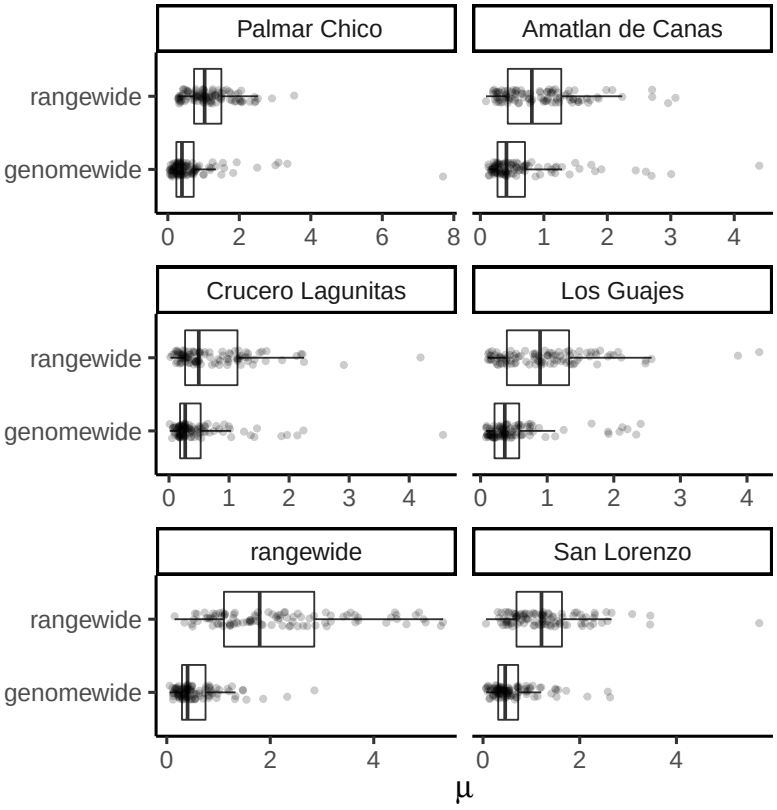

Figure S6 The distribution of RAiSD μ statistics within sweep regions unique to the rangewide sample for all maize populations. For each population, we randomly sampled 100 μ values from within the rangewide sweep regions and genomewide.