eLife

Reviewed Preprint
v2 • January 8, 2025
Revised by authors

Reviewed Preprint
v1 • November 21, 2023

Evolutionary Biology

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

Silas Tittes, Anne Lorant, Sean McGinty, James B Holland, Jose de Jesus Sánchez-González, Arun Seetharam, Maud Tenaillon, Jeffrey Ross-Ibarra

Institute of Ecology and Evolution, University of Oregon, Eugene, USA • Dept. of Evolution and Ecology, University of California, Davis, Davis, USA • Center for Population Biology, University of California, Davis, Davis, USA • Department of Plant Sciences, University of California, Davis, Davis, USA • Dept. of Integrative Genetics and Genomics, University of California, Davis, Davis, USA • United States Department of Agriculture – Agriculture Research Service, Raleigh, USA • Department of Crop and Soil Sciences, North Carolina State University, Raleigh USA • Centro Universitario de Ciencias Biológicas y Agropecuarias, Universidad de Guadalajara, Zapopan, México • Department of Ecology, Evolution, and Organismal Biology; Genome Informatics Facility, Iowa State University, Ames, USA • Génétique Quantitative et Evolution - Le Moulon, Université Paris-Saclay, INRAE, CNRS, AgroParisTech, Gif-sur-Yvette, France • Genome Center, University of California, Davis, Davis, USA

eLife Assessment

This **useful** study examines patterns of diversity and divergence in two closely related sub-species of Zea mays, patterns that have bearings on local adaptation in maize and teosinte at intermediate geographic scales. The authors suggest that convergent evolution has been facilitated by both standing variation and gene flow, with independent selective sweeps in the two species. While the data themselves are **solid**, there are limitations concerning population sampling, false positive rates in sweep detection and integration of phenotypic data, which make it difficult to draw definitive conclusions. The work should in principle be of broad interest to colleagues studying the relationship between domesticated species and their progenitors, as well as those studying instances of parallel evolution.

https://doi.org/10.7554/eLife.92405.2.sa2

Abstract

What is the genetic architecture of local adaptation and what is the geographic scale over which it operates? 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 also frequently shared by several populations, and often between 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, even after domestication, and that maize populations have facilitated adaptation in teosinte by moving beneficial alleles across the landscape. Taken together, our 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.

Introduction

As populations diverge and occupy new regions, they become locally adapted to the novel ecological conditions that they en-counter. 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 (though see Williamson *et al*. 2014; Geist *et al*. 2019 for more recent non-zero estimates of α in two plant taxa). 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 its 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. But it remains unclear if the scales identified in *Arabidopsis* are com-monly shared 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.

Using population genetic approaches, we can compare the observed distribution of beneficial alleles across multiple populations to get a more holistic description of the history of selection. 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 population genetic approaches 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 more than 9, 000 years ago (Piperno *et al*. 2009) from its common ancestor with the extant annual grass teosinte, but traditional open-pollinated populations maintain large population sizes 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; Vigouroux *et al*. 2008) and fine (Van Heerwaarden *et al*. 2010; Hufford *et al*. 2011; Takou *et al*. 2024) scales, and population genetic and common garden studies in both subspecies have shown clear signatures of populations being adapted to ecological conditions across their native range. In maize this includes local adaptation to high elevation (Gates *et al*. 2019; Janzen *et al*. 2021), 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 of local adaptation. Contrary to our predictions, our results suggest adaptations are often shared across two or more populations, and commonly between maize and teosinte. We also found that migration and standing variation have played an important role as sources 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). Both Palmar Chico

populations were down-sampled to ten randomly selected individuals to facilitate comparisons to the other populations. We did, however, use a second random sample of the Palmar Chico populations to estimate the accuracy of our inference of selective sweeps (see below in Results and Methods).

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 (**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 evidence of mixed 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 F_4 tests using Treemix (Pick-rell and Pritchard 2012) with the expectation that tests including sympatric pairs would show the strongest signatures of gene flow. Instead, we found that all populations showed some evidence of gene flow with various populations of the other sub-species, as measured by the high absolute Z-Scores of the F_4 statistic. We found little evidence of increased gene flow between sympatric pairs (**Figure 1E**), but Z-Scores were sensitive to the specific combinations of non-focal populations included in each test (**Table S2**). Specifically, we found that elevated F_4 tests almost always included the maize population from Crucero Lagunitas ($p < 2 \times 10^{-10}$), which was true whether or not the F_4 test included its sympatric teosinte. Finding higher evidence of gene flow when maize from Crucero Lagunitas is included in the test is interesting in combination with results from selective sweeps, where it also plays an outsized role (see below).

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 (**Figures 2 A** and **B**, **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**).

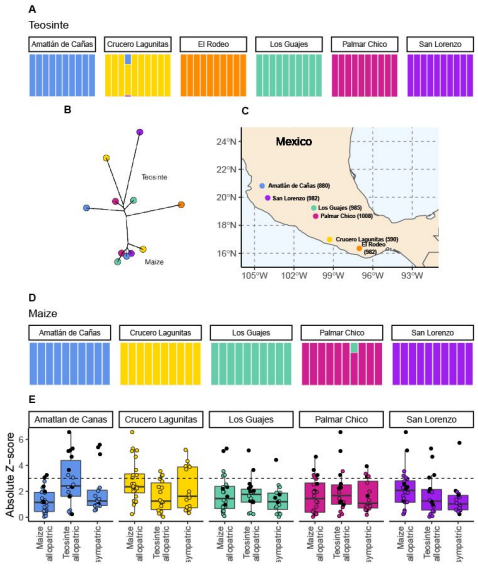


Figure 1.

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) F_4 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 F_4 test, where a given focal population was partnered with another population of the same subspecies as a sister node, and two other populations from the other sub-species as a sister clade (see Methods for further details). Black points show F_4 tests that included maize from Crucero Lagunitas, otherwise points are colored by focal population. The dotted line corresponds to our chosen significance threshold (p = 0.001).

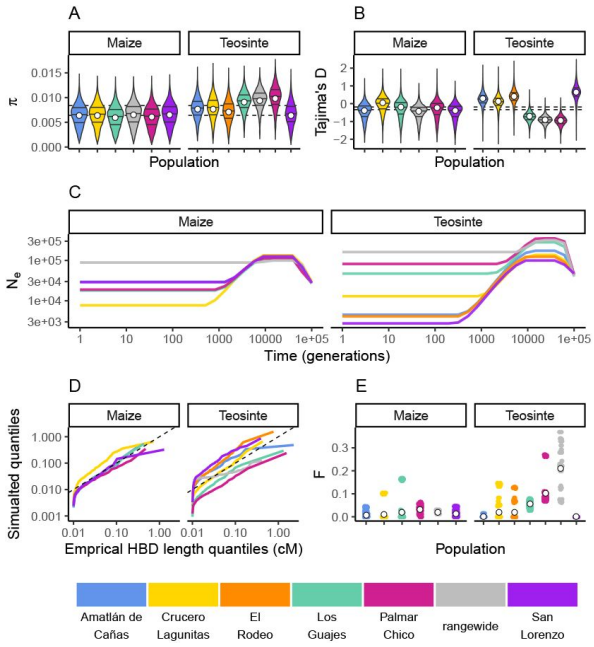


Figure 2.

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 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 in each population. Filled white points are the average values for each population.

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 Brown-ing 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 may not be as accurately captured by 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 inbreeding coefficients (F) using ngsRelate (Hanghøj *et al*. 2019). Although inbreeding coefficients were as high as 0.37, the mean value of F was 0.017 ± 0.001 (SE) and 0.033 ± 0.001 (SE) for maize and teosinte (respectively), (**Figure 2E**). These values are consistent with prior estimates of the rate of outcrossing of ≈ 3% in teosinte (Hufford *et al*. 2011) and suggest there has been relatively little inbreeding in either subspecies in the recent past.

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. This pattern should become even stronger if alleles experience negative pleiotropy, such that they are only adaptive in one environment and deleterious in others, further inhibiting the ability of such alleles to fix rangewide. 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 mixedeffect 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 3**). 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 S2**, Supplement III), 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.

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 any-where in the genome, we used RAiSD (Alachiotis and Pavlidis 2018) to

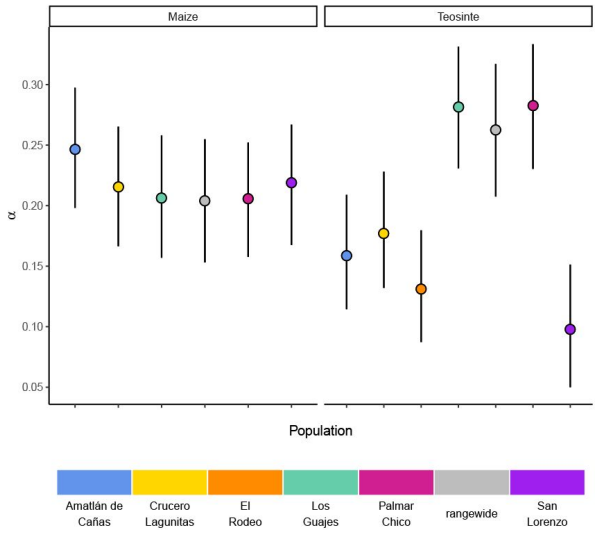


Figure 3.

The proportion of mutations fixed by natural selection.

Estimated values of the proportion of mutations fixed by natural selection (α) by population. Vertical lines show the 95% credible interval.

identify putative selective sweeps in each population, where sweep regions were identified by merging μ summary statistic outliers using a threshold defined by coalescent simulations under each population's estimated demography (see Methods). Simulations suggest this approach has high precision and power compared to alternative methods over a broad range of 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. After optimizing over a grid of hyperparameters for RAiSD (see Methods), we estimated the proportion of sweeps shared between replicates to be 0.67 and 0.80 for maize and teosinte, respectively. This proportion is consistent with the false positive rates for strongly selected hard sweeps estimated from simulations we conducted under the demographic histories (See Supplement IV). As such, a non-trivial number of the putative sweep regions we inferred with RAiSD in other populations are likely also false positives, though far fewer than previously reported using alternative methods for identifying sweep regions (Tittes *et al*. 2021).

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 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**). Across teosinte populations, 22% of sweeps were private, which was significantly greater than the 14% found in maize (binomial glm, p = 0.0026; **Figure 4B**). These proportions could be an overes-timate because of relatively high false negative rates, causing us to count shared sweeps as unique (See **Table S4**). This likely does not impact the difference in proportions between species, however, since the average power to detect sweeps is similar in both subspecies (0.67 and 0.8 in maize and teosinte, respectively; see Methods).

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 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 our estimate of sweep accruacy (see Methods). Sympatric pairs did not tend to have a lower p value than allopatric pairs, and no population pair 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 4D**), 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 101 sweeps that were shared by two or more populations, there were 1, 50, and 40 sweeps inferred to be convergent via independent mutations, migration, or standing variation; and an additional 10 sweeps inferred to be neutral (**Figure 5C**). 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**). Further, across sites proposed to be the locus of the beneficial mutation there was a strong tendency for the

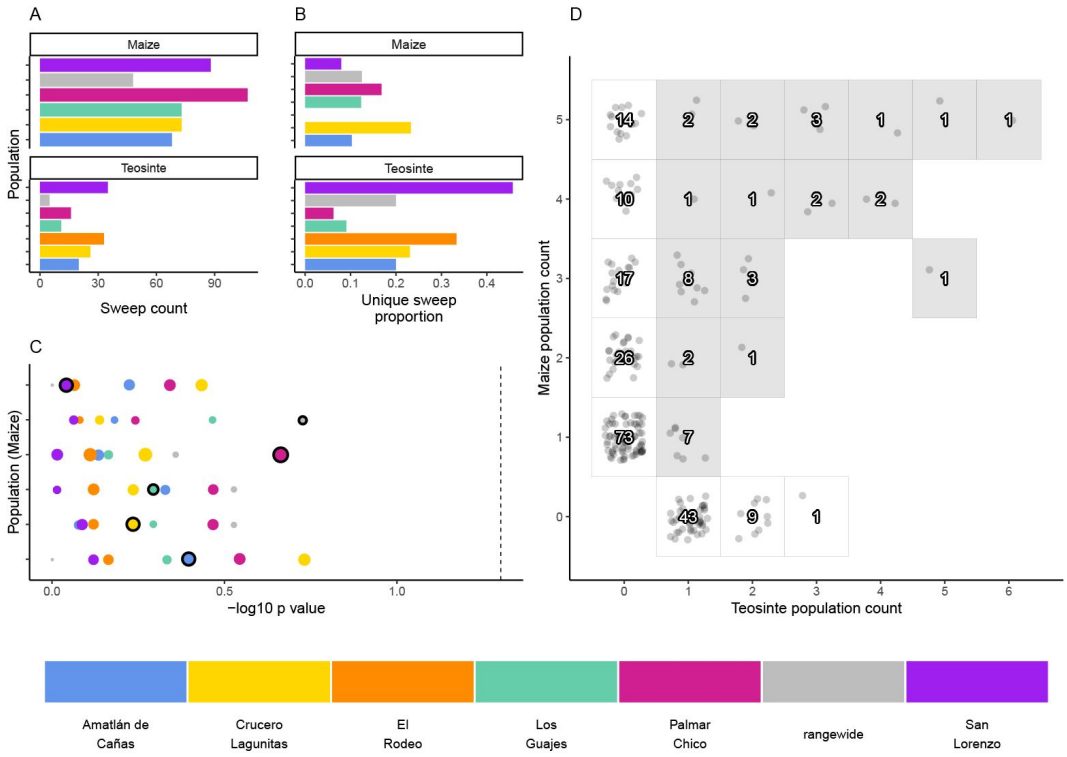


Figure 4.

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. (C) Negative $\log_10$ 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 at the bottom), while the point color designates the teosinte population each maize population was paired with. Points with black outline 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) 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 sub-species.

next best composite likeli-hood score to be at a site nearest to the maximum composite likelihood site, suggesting relatively high confidence in the location of the sweep (**Figure S8**). 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 sub-species, no one population stood out as the most frequent, with fairly even counts across populations of both subspecies. In contrast, convergence via migration exclusive to teosinte was seen rarely and from only three source populations, while the most common source population for convergence via migration exlusive to maize was almost entirely from Crucero lagunitas (**Figure 5D**). We provide hypotheses for why maize from Crucero lagunitas would be a common source population for beneficial alleles in maize only, and how this relates to our F_4 tests findings (**Figure 1**) in the discussion.

In convergence models with migration, we tested migration rates between 10^{-3} and 10^{-1}. The most likely migration rate varied across sweeps, but tended to be 10^{-1} for the majority of sweeps shared by the two subspecies and sweeps exclusive to maize. In contrast, the lowest migration rate (10^{-3}) was al-ways the most likely for sweeps exclusive to teosinte (**Figure 5F**), although there were only five instances total, suggesting convergence via migration exclusive to teosinte at any rate is rare. 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, but that distinct patterns emerge in the presence or absence of maize populations sharing the signature of adaptation.

Discussion

Local adaptation occurs at intermediate scales

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 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 S2**). However, we acknowledge that our evidence relies primarily on the patterns found in teosinte; due to a shared domestication history maize populations have less independent evolution and a larger proportion of fixed differences are shared among all populations, leading to a similar α estimate between individual populations and the rangewide sample. Overall our findings are 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 and antagonistic pleiotropy suppresses rangewide fixation in alternative environments. Our findings make sense in the light of other work studying pleiotropy's impact on adaptive evolution in *Arabidopsis*, which found that most mutations impact few traits and that the genetic architecture was largely non-overlapping when studied across multiple environments (Frachon *et al*. 2017). Our results are consistent with models in which locally beneficial alleles are simply neutral elsewhere, a process thought to be more common when overall levels of gene flow are low (Wadgymar *et al*. 2022). One possible explanation for the considerable overlap in α between rangewide and individual populations is that most of the fixations occur on long branches shared by all populations. This does not appear to be the case (see **Table S3**), likely because the coalescent within *Zea mays* ssp. *parviglumis* is a substantial fraction of the relatively recent divergence to the outgroups *Z. diploperennis* and *Z. luxurians* (Chen *et al*. 2022). Lastly, it is worth nothing that while the asymptotic MK method we employed has been shown to provide reliable estimates of α when fixations are due to strong beneficial

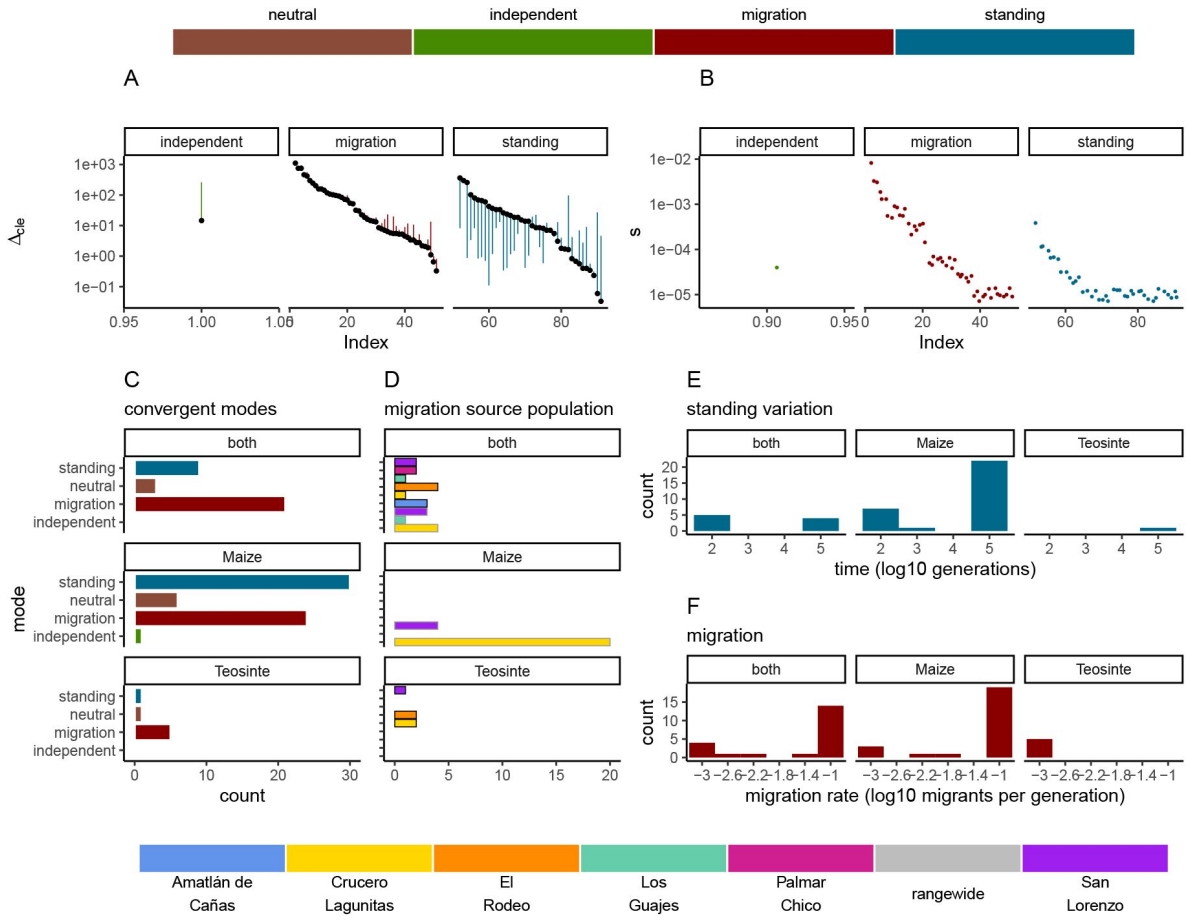


Figure 5.

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.

mutations (Messer and Petrov 2013), it assumes there is no contribution of positively selected polymorphisms, and as such, our estimates may be conservative compared to methods that account for them (Uricchio *et al*. 2019).

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, no pairs 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**). This 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, yet some adaptive teosinte introgression appears widespread among highland maize.

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, are best-suited for adaptation that leads to fixation of beneficial alleles, and/or mutations of large effect. For the moderate population sizes and selection coefficients observed here, fixation of new beneficial mutations takes a consid-erable amount of time, on the order of $4log(2N)/s$ generations (Charlesworth 2020). 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 may have been under selection in ancestral populations that occupied different environments than the sampled individuals, and their signatures may no longer be detectable (Przeworski 2002), placing a limit on the temporal resolution with which we can make inference about instances of local adaptation. Conversely, it is possible that some of our shared sweeps represent selection that happened prior to divergence between closely related populations. This could explain why some populations (teosinte from Los Guajes and Palmar Chico) have fewer unique sweeps (**Figure 4B**), and would similarly create a mismatch in the conditions that populations currently occupy and those of the ancestral population where selection was experienced. 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 also 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 genetic basis of many adaptive traits may be highly polygenic. Allele frequency changes underlying poly-genic 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, and especially if it were acting on alleles with small effect sizes, we would expect to find few to no shared sweeps. The fact that we find many instances of sharing across populations suggests 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 expected under a selective sweep model.

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**). 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 sometime prior to ≈ 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 domestication, it is less clear why we found a similar 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, Mesoamerican phytolith records in lake sediments show evidence of anthropogenic burning as early as 11K years B.P. (Piperno 1991). The establishment and spread of human populations over the subsequent millennia would require an 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 specieswide 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 with approaches based on the site frequency spectrum (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.

Demographic inferences are known to be sensitive to the effects of linked selection (Ewing and Jensen 2016). 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, which has had recent explosive population growth, are occurring in the more recent past (Beissinger *et al*. 2016).

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 4**). 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 some combination of longer divergence times, reduced gene flow, and/or higher genetic drift from reduced population sizes

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*. 2021) 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.

For convergent sweeps that occurred via standing variation within maize or 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 **Figure 5E**). In contrast, the sweep exclusive to teosinte was inferred to be standing variation for more generations, although there was only one such instance. Sweeps that occurred via standing variation and shared between sub-species 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 bimodal features of sharing seems at face value surprising. How can an allele be standing variation for so many generations after divergence but prior to selection when the populations diverged less than ten thousand generations ago? We speculate that ancestral population structure and limited sampling of 6 populations could explain the pattern. For example, extremely long standing times that predate domestication may reflect divergence between our sampled teosinte populations and the populations most closely related to those that gave rise to domesticated maize. More comprehensive sampling could help to resolve patterns of local adaption.

The most common mode of convergent adaptation was via migration, and frequently occurred between geographically disparate populations (**Figures S6**). This included a relatively large number of shared sweeps via migration between maize and teosinte (**Figures 5** and **S7**). 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), and previous work has identified gene flow between geographically disparate populations of 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*. 2021).

Our findings on convergence via migration point to an intriguing hypothesis, namely, that some number of alleles that are beneficial to teosinte may have originally arisen in a different teosinte or maize 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. It is worth noting that the total number of shared sweeps is much lower than previously reported (Tittes *et al*. 2021), and only four of fifteen instances of convergence via migration include two or more teosinte and

maize populations (**Figure S7**). Despite the lower totals, the evidence still supports this hypothesis. 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 (via human movement and trade). 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 strong evidence. Further, sweeps shared via migration that were exclusive to teosinte were always inferred to be the lowest migration rate (1×10^{-3}) (**Figure 5F**), where those shared by both subspecies or exclusive to maize were primarily inferred to be the highest 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 sub-species, there is evidence of gene flow based on F_4 tests. In particular, F_4 tests that included maize from Crucero Lagunitas were consistently elevated across both subspecies (**Figure 1E** and **S1**). It was surprising to find that maize from Crucero Lagunitas was featured heavily as a source population, but only for sweeps exclusive to maize (**Figure 5D**). One explanation for this discrepancy is that perhaps movement of this population was common, but the alleles were only beneficial in a domesticated genetic background and have been excluded from teosinte populations except in the case of neutral regions, which could leave a signature in F_4 tests but not tests for beneficial mutations. 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 often the source of beneficial alleles for sweeps shared between both subspecies (**Figure 5C**). As such, teosinte populations may have also commonly been the source of adaptive variation for both other teosinte and maize populations.

Materials and methods

Samples and whole genome resequencing

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 which 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 a Qubit (Life Technologies) and 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 to an average coverage of 20 to 25x PE150 on the Illumina X10 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 BioProject ID PRJNA1076902.

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, 100kb and 1,000kb windows with the *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 proce-dure 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 F_4 statistics and Z-scores calculated with blocked jackknifing, implemented using treemix (Pickrell and Pritchard 2012). Trees for F_4 tests were always of the form (*Maize_X, Maize_Y*; *Teosinte_ f ocal, Teosinte_Z*), or (*Maize_ f ocal, Maize_Y*; *Teosinte_W, Teosinte_Z*); with each unique combination of populations considered to be the "*f ocal*" 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 covariance 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.

Estimating the Proportion of Fixations Due to Positive Selection, α

We modeled the rate of positive selection, α of 0-fold nonsynonymous 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 $(n-1)/n$) (see section "Genetic Diversity" above for further details on uSFS estimation). 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, respectively, 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 i^{th} allele frequency bin of the j^{th} 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 III), 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 signatures of selective sweeps 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 24 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 $1 \times 10_{-6}$. 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).

Individual SNP windows in RAiSD are small and spatially auto-correlated with neighboring windows, implying that a sweep signature is distributed over multiple windows. This makes it difficult to compare shared RAiSD outliers across populations, which may have experienced the same sweep, but may differ in how the μ statistic is distributed across windows. To compare sweep signatures across populations we collected outliers into "sweep regions" by fitting a cubic spline modeling the adjusted μ statistic by position along the chromosome using the *smooth.spline* function in R, which uses cross validation to choose the optimal smoothing parameter for the cubic spline. We compared the empirical cubic spline fit to the those simulated under each population's demography using msprime (Kelleher *et al*. 2016), and selected windows with fitted values greater than 99.9th percentile from the simulations as outlier regions. We merged outlier regions within 50 Kb of one another and treated them as a single sweep region. The custom R functions for defining outlier regions are available at *https://github.com/silastittes/sweep_regions*.

For the above steps, we selected the best combination of parameters by conducting a grid search that optimized on the highest proportion of sweeps shared between the two Palmar Chico replicates, averaged over both subspecies. For the grid search we considered RAiSD snp windows of 10, 24, 50, 100, 200, and 500; outlier quantiles of 0.8, 0.9, 0.95, 0.99, and 0.999; and merged windows of 50K, 100K, 200K, and 500K BPs.

Assessing the accuracy of sweep inferences and the number of shared sweeps expected by chance

To assess the accuracy of sweep inferences, we used a second random sample of 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

$$P = 1 - \left(\frac{n_S}{n_{P1}} + \frac{n_S}{n_{P2}}\right)/2$$

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 P for the two subspecies, although it was higher for maize than for teosinte (0.33 and 0.20, 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 sweep inference accruacy by multiplying the raw values of N, n_1, n_2, and x by the value of P 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 of 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 similar densities, with a lower and upper bound of 1K and 100K SNPS, respectively. 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 to neutral expectations, we first sampled allele frequencies at 100K random loci. When fitting *rdmc*, we assumed the effective population size was 50K for all populations. 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 includes arguments that control the grid of parameter values over which composite likelihoods were computed. These were:

n_sites = 200, num_bins = 1000, sels = 10^seq(-5, -1, length.out = 12), times = c(1e2, 1e3, 1e4, 1e5), gs = 10^seq(-3, -1, length.out = 6), migs = 10^(seq(-3, -1, length.out = 6)), 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 suite (Wickham *et al*. 2019). Figures were made using ggplot2 (Wickham 2016), patchwork (Pedersen 2019), and cowplot (Wilke 2019). Code and instructions for the entirety of the analyses, including Jupyter notebooks for reproducing figures, is available from *https://github.com/silastittes/parv_local*.

Acknowledgements

We would like to thank Andi Kur for providing the corn art, along with Matthew Gibson, Tom Booker, Cathy Rushworth, and 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) and Hatch support from the USDA (project CA-D-PLS-2066-H 548). We would also like to acknowledge Felix Andrews for statistical advice, although we did not follow it.

Appendix

Supplement I Population sampling locations

Supplement II Further assessment of f4 statistic inferences

Supplement III Additional analyses related to α

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.

Amount of unique fixations across teosinte populations

Supplement IV Further assessment of sweep inferences and precision

We found that only 67% and 80% of sweeps were shared between the two random subsamples for maize and teosinte populations from Palmar Chico, respectively (see Results). While our updated method to identifying sweeps improved precision over our previous one (which shared 40% and 50% (Tittes *et al.* 2021)), the low precision still warrants further exploration.

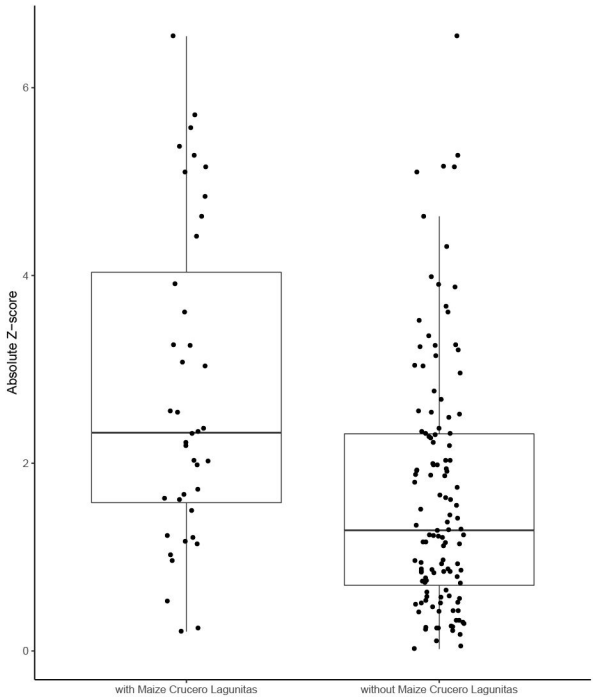
Table S1.

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

Figure S1.

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



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

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.

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

Table S2. (continued)

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

Table S2. (continued)

Figure S2.

Predicted values of *α* across mutation types.

Grey bands for each mutation type show the 95% credible intervals averaged over each population.

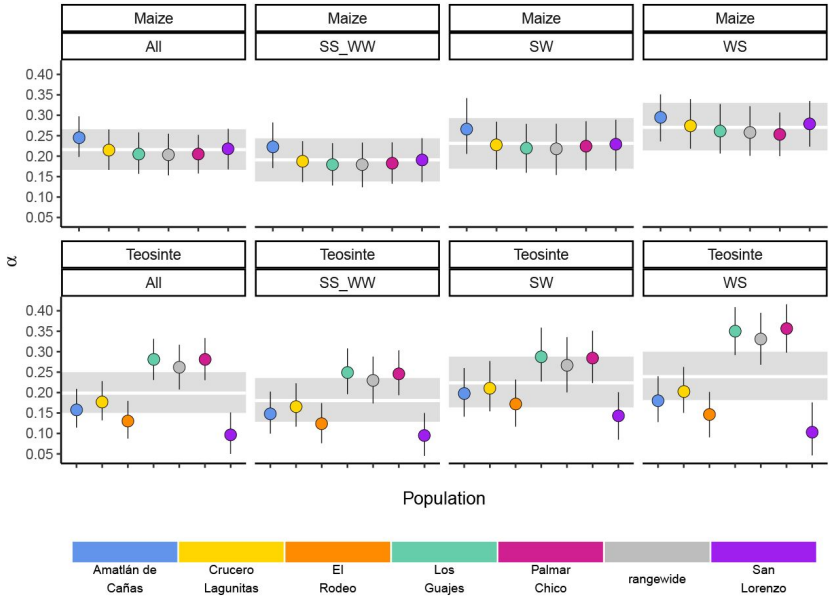


Table S3.

Proportion and count of fixed differences at 0 fold sites across teosinte populations

Population	proportion	total
Amatlan de Canas	.12238	1195
Crucero Lagunitas	.15619	1880
El Rodeo	.23233	2983
Los Guajes	.03025	170
San Lorenzo	.33796	6050
Palmar Chico	.03580	184

One explanation for the low sweep precision could be substructure within the Palmar Chico populations, leading to replicates with slightly different histories. This could create unequal power to detect sweeps if, for example, the subpopulations had different progress towards fixation of the beneficial allele. However, this explanation is unlikely as individuals were randomly assigned to subpopulations, so any substructure is likely to be distributed evenly between the two samples. This is further supported from the two samples showing relatively short branch lengths in the population phylogeny (**Figure S3**). The branch lengths separating the two subsamples (cophenetic distance) was 0.00835 and 0.00962 for maize and teosinte respectively, compared to the within subspecies means of 0.0170 and 0.0559.

Another potential explanation for lowered sweep sharing between replicates is that sweeps vary in their detectability based on their characteristics. Namely, sweeps that were weakly selected, incomplete, and/or ones that started at a high initial frequency prior to the onset of selection (soft sweeps) may vary in their detectability using the methods we employed. We conducted a simulation experiment to better understand the potential causes of the low shared proportion, and to measure performance to detect different kinds of sweeps more generally. We used discoal (Kern and Schrider 2016) to simulate sweeps in a 400Kb region using the average genome-wide maize mutation and recombination rates under the inferred demographic history of the maize population from Palmar Chico (**Figure 2**). We simulated four distinct scenarios: classical hard sweeps, where selection acts to fix an an adaptive mutation; soft sweeps, where selection is initiated after the adaptive allele reaches a specified frequency; and incomplete sweeps, where a hard sweep simulation is stopped at a specified frequency, and neutral simulations without selection. For soft sweeps, we varied the initial by drawing from a beta distribution with shape parameters 1 and 20 . Incomplete sweeps finished when the adaptive allele reached a frequency of 0.5. For all three types of sweeps, we also varied the strength of selection using the parameter $\alpha = 4N0s$ to be 10, 50, or 100, where $N0$ is the present day effective population size and s is the selection coefficient. In addition to matching demography and other parameters, we used the same sampling scheme, simulation 50 individuals, than randomly choosing two non-overlapping subsets of 10 individuals (*https://github.com/silastittes/ms_sub*). From the simulations we assessed the True/False Positive/Negative Rates for each combination of sweep type and strength of selection (α), as well as the distribution of base pair overlap between sweep regions inferred in the the two random subsamples. The same sweep inference methods and parameters were used for these simulations and the empirical samples (see methods). Overall, we found that sweep characteristics we explore indeed impacted our overall power to detect them, and the about of overlap between the sweep regions. Namely, weakly selected sweeps had consistently lower True Positive Rates (**Table S4**).

Figure S3.

Treemix phylogeny including both subsamples of Palmar Chico.

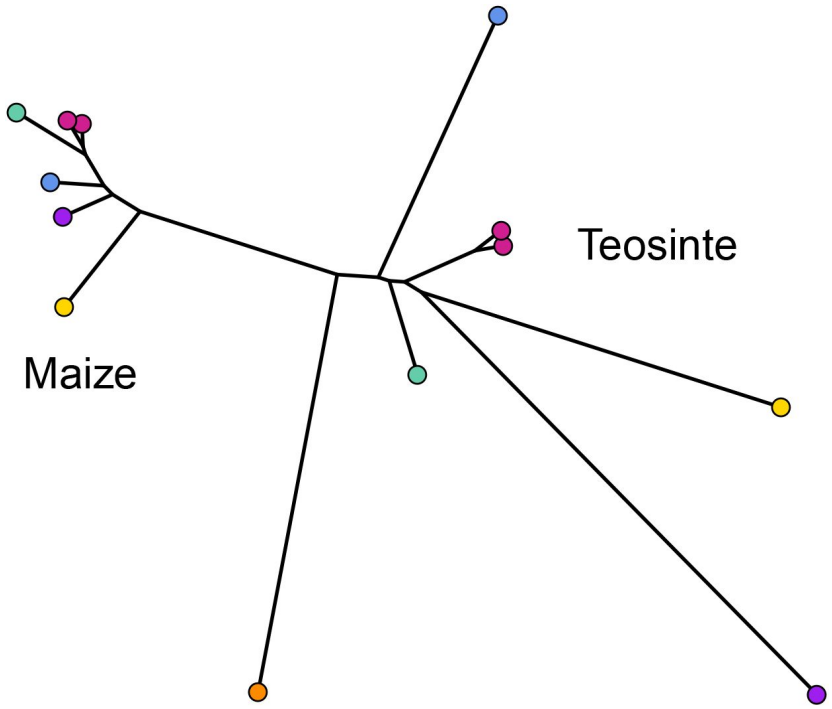


Table S4.

Performance to detect simulated hard, soft, and incomplete sweeps under varying strengths of selection under the maize Palmar Chico population demography.

TPR, TNR, FNR, and FPR stand for true postive, true negative, false negative, and false positive rates, respectively.

$4N_es$	simulation type	TPR	TNR	FNR	FPR
10	hard	0.14	0.67	0.86	0.33
50	hard	0.92	0.75	0.08	0.25
100	hard	0.99	0.84	0.01	0.16
10	incomplete	0.04	0.67	0.96	0.33
50	incomplete	0.06	0.67	0.94	0.33
100	incomplete	0.05	0.63	0.95	0.37
10	soft	0.13	0.71	0.87	0.29
50	soft	0.75	0.74	0.25	0.26
100	soft	0.81	0.76	0.19	0.24

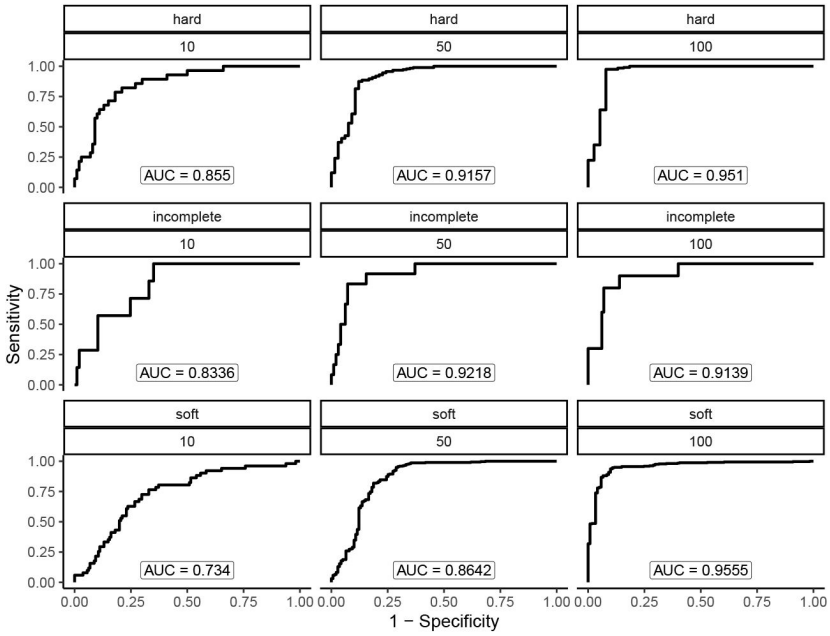


Figure S4.

Performance to detect simulated hard, soft, and incomplete sweeps under varying strengths of selection under the maize Palmar Chico population demography.

Each panel shows a combinations of sweep type (hard, soft, or incomplete) and strength of selection ($\alpha = 4N_e s$ = 10, 50, or 100)

Figure S5.

Degree of overlap between simulated sweep regions takenfrom two downsampled replicates under the maize Palmar Chico population demography.

Positive values show the amount of overlap in basepairs between sweep regions, while negative values represent a the space between them. Panel structure follows that of S4.

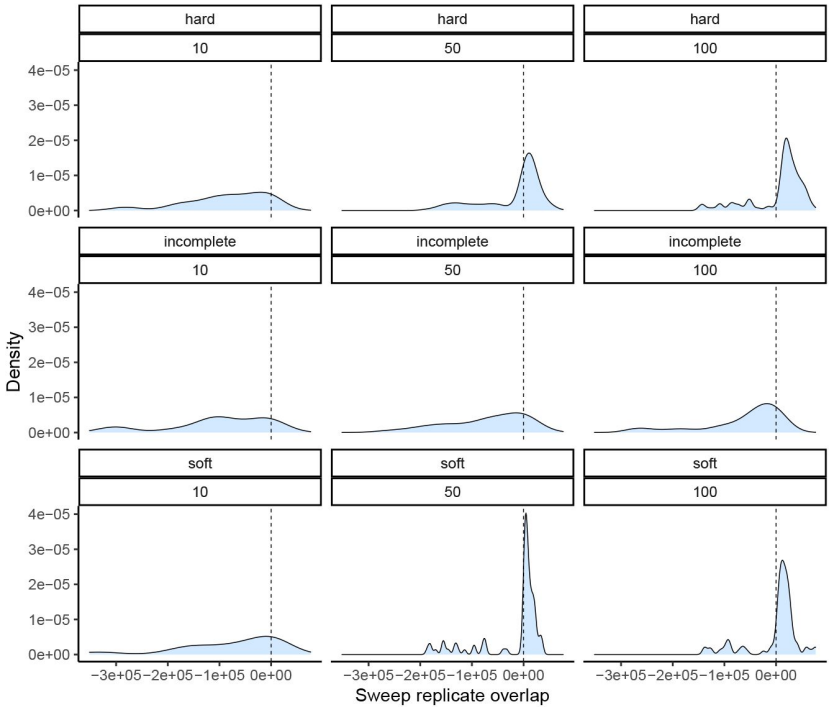


Figure S6.

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.

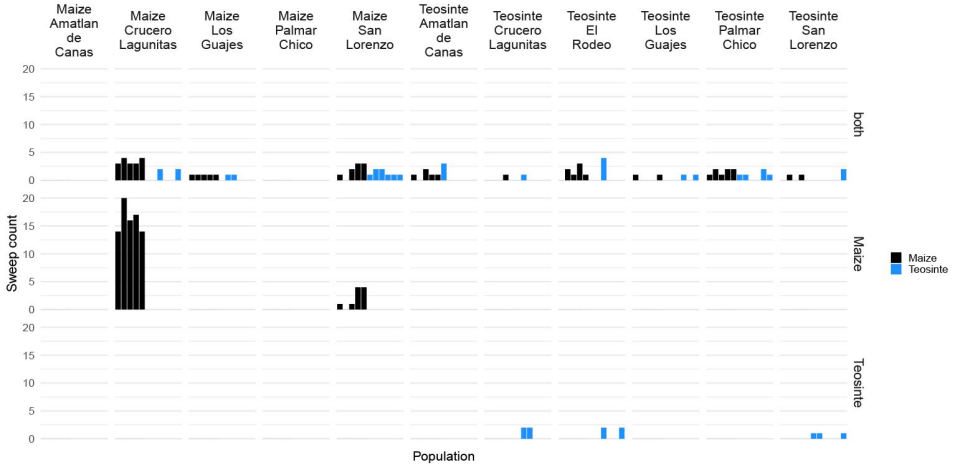


Figure S7.

Inferred sweeps shared between subspecies 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.

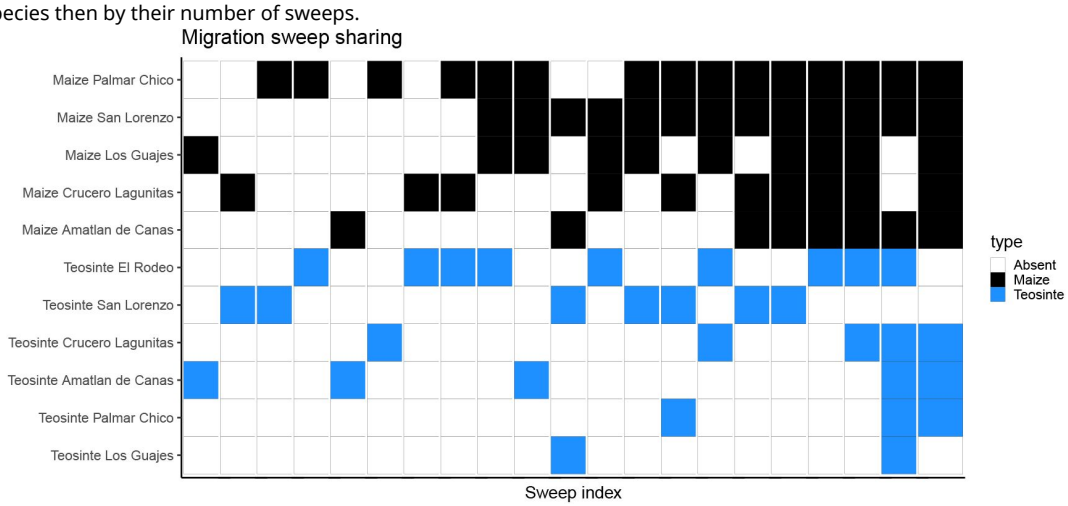
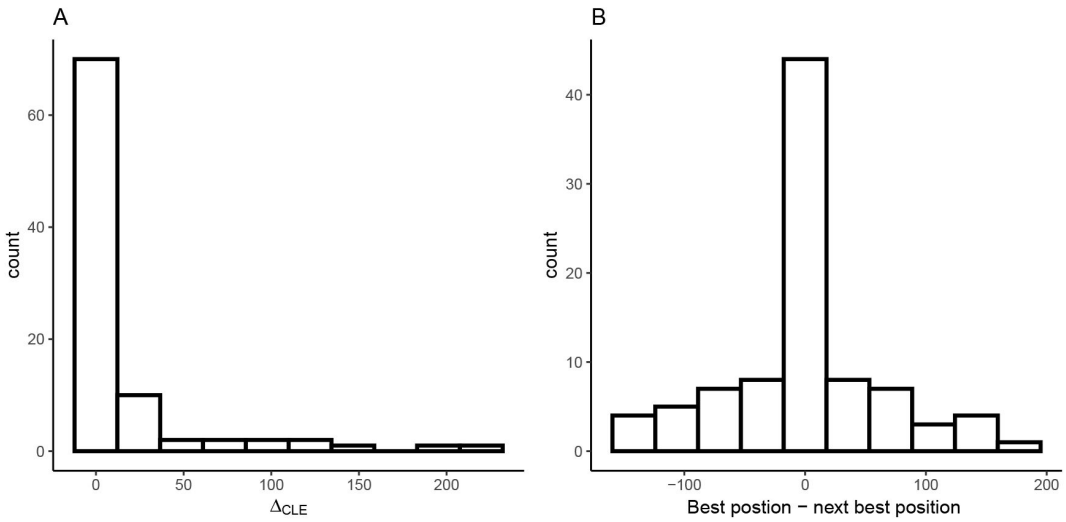


Figure S8.

Variation in composite likelihoods within sweep regions.

(A) Distribution of differences between highest and next highest composite likelihoods each sweep region. (B) Distribution of differences between highest and next highest composite likelihoods candidate beneficial mutation positions each sweep region.



References

Aguirre-Liguori J. A., Gaut B. S., Jaramillo-Correa J. P., Tenaillon M. I., Montes-Hernández S., 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., Pavlidis P. (2018) **Raisd detects positive selection based on multiple signatures of a selective sweep and snp vectors** *Communications biology* **1**:1–11

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

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

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

Bourne E. C., Bocedi G., Travis J. M., Pakeman R. J., Brooker R. W., 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**

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

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

Butler E. E., Huybers P. (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., Gates D., Lorant A., Perkins M. T., Coop G., 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 L., Luo J., Jin M., Yang N., Liu X., et al. (2022) **Genome sequencing reveals evidence of adaptive variation in the genus zea** *Nature Genetics* **54**:1736–1745

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

Chevin L.-M., Martin G., Lenormand T. (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., Tavaré S., Doebley J. (2005) **Estimating a nucleotide substitution rate for maize from polymorphism at a major domestication locus** *Molecular biology and evolution* **22**:2304–2312

Clausen J., Keck D. D., Hiesey W. M., et al. (1948) **Experimental studies on the nature of species. III. Environresponses of climatic races of Achillea** Carnegie Institution of Washington, Washington, DC

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

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

DeWitt W. S., Harris K. D., Ragsdale A. P., Harris K. (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** *Isozymes in plant biology* Springer :165–191

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

Ewing G. B., Jensen J. D. (2016) **The consequences of not accounting for background selection in demographic inference** *Molecular ecology* **25**:135–141

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

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

Frachon L., Libourel C., Villoutreix R., Carrère S., Glorieux C., et al. (2017) **Intermediate degrees of synergistic pleiotropy drive adaptive evolution in ecological time** *Nature ecology & evolution* **1**:1551–1561

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

Fustier M.-A., Brandenburg J.-T., Boitard S., Lapeyronnie J., Eguiarte L., 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., Martínez-Ainsworth N.E., Aguirre-Liguori J. A., Venon A., Corti H., et al. (2019) **Common gardens in teosintes reveal the establishment of a syndrome of adaptation to altitude** *PLoS genetics* **15**

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

Geist K. S., Strassmann J. E., Queller D. C. (2019) **Family quarrels in seeds and rapid adaptive evolution in arabidopsis** *Proceedings of the National Academy of Sciences* **116**:9463–9468

Gossmann T. I., Song B.-H., Windsor A. J., Mitchell-Olds T., Dixon C. J., 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., Haines M. R., Steckel R. H. (2000) **A population history of North America** Cambridge University Press

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

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

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

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., Gepts P., ROSS-IBARRA J. (2011) **Influence of cryptic population structure on observed mating patterns in the wild progenitor of maize (zea mays ssp** *parviglumis). Molecular ecology* **20**:46–55

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

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

Janzen G. M., Aguilar-Rangel M. R., Cíntora-Martínez C., Blöcher-Juárez K.A., González-Segovia E., et al. (2021) **Demonstration of local adaptation of maize landraces by reciprocal transplantation** *bioRxiv*

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

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

Kern A. D., Schrider D. R. (2016) **Discoal: flexible coalescent simulations with selection** *Bioinformatics* **32**:3839–3841

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

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

Li H. (2013) **Aligning sequence reads, clone sequences and assembly contigs with bwa-mem** *arXiv*

Lowry D. B., Rockwood R. C., Willis J. H. (2008) **Ecological reproductive isolation of coast and inland races of mimulus guttatus** *Evolution: International Journal of Organic Evolution* **62**:2196–2214

Mei W., Stetter M. G., Gates D. J., Stitzer M. C., Ross-Ibarra J. (2018) **Adaptation in plant genomes: Bigger is different** *American Journal of Botany* **105**:16–19

Messer P. W., Petrov D. A. (2013) **Frequent adaptation and the mcdonald–kreitman test** *Proceedings of the National Academy of Sciences* **110**:8615–8620

Nannas N. J., Dawe R. K. (2015) **Genetic and genomic toolbox of zea mays** *Genetics* **199**:655–669

O'Brien A. M., Sawers R. J., Strauss S. Y., Ross-Ibarra J. (2019) **Adaptive phenotypic divergence in an annual grass differs across biotic contexts** *Evolution* **73**:2230–2246

Ogut F., Bian Y., Bradbury P. J., Holland J. B. (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., Perales H., Hijmans R. J. (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., Schneeberger K., Lucas-Lledó J.I., Warthmann N., Clark R. M., 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** *CRAN*

Pickrell J., Pritchard J. (2012) **Inference of population splits and mixtures from genome-wide allele frequency data** *Nature Precedings* :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., Ranere A. J., Holst I., Iriarte J., Dickau R. (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., Woodhouse M. R., Cannon E. K., Gardiner J. M., Harper L. C., et al. (2019) **Maizegdb 2018: the maize multi-genome genetics and genomics database** *Nucleic acids research* **47**:D1146–D1154

Pritchard J. K., Pickrell J. K., Coop G. (2010) **The genetics of human adaptation: hard sweeps, soft sweeps, and polygenic adaptation** *Current biology* **20**:R208–R215

Przeworski M. (2002) **The signature of positive selection at randomly chosen loci** *Genetics* **160**:1179–1189

Pyhäjärvi T., Hufford M. B., Mezmouk S., Ross-Ibarra J. (2013) **Complex patterns of local adaptation in teosinte** *Genome biology and evolution* **5**:1594–1609

R Core Team (2020) **R: A Language and Environment for Statistical Computing** Vienna, Austria: R Foundation for Statistical Computing

Ranum P., Peña-Rosas J.P., Garcia-Casal M. N. (2014) **Global maize production, utilization, and consumption** *Annals of the new York academy of sciences* **1312**:105–112

Rodriguez-Zapata F., Barnes A. C., Blocher-Juarez K. A., Gates D. J., Kur A., et al. (2021) **Teosinte introgression modulates phosphatidylcholine levels and induces early maize flowering time** *bioRxiv*

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

Rudman S. M., Greenblum S. I., Rajpurohit S., Betancourt N. J., Hanna J., et al. (2021) **Direct observation of adaptive tracking on ecological timescales in drosophila** *bioRxiv*

Savolainen O., Lascoux M. (2013) **Ecological genomics of local adaptation** *Nature Reviews Genetics* **14**:807–820

Schrider D. R., Kern A. D. (2016) **S/hic: robust identification of soft and hard sweeps using machine learning** *PLoS genetics* **12**

Skotte L., Korneliussen T. S., Albrechtsen A. (2013) **Estimating individual admixture proportions from next generation sequencing data** *Genetics* **195**:693–702

Smith N. G., Eyre-Walker A. (2002) **Adaptive protein evolution in drosophila** *Nature* **415**:1022–1024

Stitzer M. C., Ross-Ibarra J. (2018) **Maize domestication and gene interaction** *New phytologist* **220**:395–408

Swarts K., Gutaker R. M., Benz B., Blake M., Bukowski R., et al. (2017) **Genomic estimation of complex traits reveals ancient maize adaptation to temperate north america** *Science* **357**:512–515

Takou M., Schulz K., Stetter M. G. (2024) **Local selection shaped the diversity of european maize landraces** *bioRxiv* :2024–6

Tenaillon M. I., Hufford M. B., Gaut B. S., Ross-Ibarra J. (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

Tittes S., Lorant A., McGinty S., Doebley J. F., Holland J. B., et al. (2021) **Not so local: the population genetics of convergent adaptation in maize and teosinte (version 1)** *bioRxiv*

Tuanmu M.-N., Jetz W. (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., Petrov D. A., Enard D. (2019) **Exploiting selection at linked sites to infer the rate and strength of adaptation** *Nature ecology & evolution* **3**:977–984

Van Heerwaarden J., Ross-Ibarra J., Doebley J., Glaubitz J. C., De Jesús Sánche González J., et al. (2010) **Fine scale genetic structure in the wild ancestor of maize (zea mays ssp. parviglumis)** *Molecular Ecology* **19**:1162–1173

Vigouroux Y., Glaubitz J. C., Matsuoka Y., Goodman M. M., Sánchezg J., et al. (2008) **Population structure and genetic diversity of new world maize races assessed by dna microsatellites** *American journal of botany* **95**:1240–1253

Wadgymar S. M., DeMarche M. L., Josephs E. B., Sheth S. N., Anderson J. T. (2022) **Local adaptation: Causal agents of selection and adaptive trait divergence** *Annual Review of Ecology, Evolution, and Systematics* **53**:87–111

Wang L., Beissinger T. M., Lorant A., Ross-Ibarra C., Ross-Ibarra J., et al. (2017) **The interplay of demography and selection during maize domestication and expansion** *Genome biology* **18**:1–13

Wang L., Josephs E. B., Lee K. M., Roberts L. M., Rellán-Álvarez R., et al. (2021) **Molecular parallelism underlies convergent highland adaptation of maize landraces** *Molecular biology and evolution* **38**:3567–3580

Whitehead A., Galvez F., Zhang S., Williams L. M., Oleksiak M. F. (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** New York: Springer-Verlag

Wickham H., Averick M., Bryan J., Chang W., McGowan L. D., et al. (2019) **Welcome to the tidyverse** *Journal of Open Source Software* **4**

Wilke C. O. (2019) **cowplot: Streamlined Plot Theme and Plot Annotations for 'ggplot2'** *CRAN*

Wilkes H. G., et al. (1967) **Teosinte: the closest relative of maize** Bussey Institution of Harvard University

Williamson R. J., Josephs E. B., Platts A. E., Hazzouri K. M., Haudry A., et al. (2014) **Evidence for widespread positive and negative selection in coding and conserved noncoding regions of capsella grandiflora** *PLoS genetics* **10**

Wright S. I., Bi I. V., Schroeder S. G., Yamasaki M., Doebley J. F., et al. (2005) **The effects of artificial selection on the maize genome** *Science* **308**:1310–1314

Xue A. T., Schrider D. R., Kern A. D. (2021) **Discovery of ongoing selective sweeps within anopheles mosquito popula-populations using deep learning** *Molecular biology and evolution* **38**:1168–1183

Author information

Silas Tittes

Institute of Ecology and Evolution, University of Oregon, Eugene, USA, Dept. of Evolution and Ecology, University of California, Davis, Davis, USA, Center for Population Biology, University of California, Davis, Davis, USA
ORCID iD: 0000-0003-4697-7434

For correspondence: stittes@ucdavis.edu

Anne Lorant

Department of Plant Sciences, University of California, Davis, Davis, USA

Sean McGinty

Dept. of Integrative Genetics and Genomics, University of California, Davis, Davis, USA

James B Holland

United States Department of Agriculture – Agriculture Research Service, Raleigh, USA, Department of Crop and Soil Sciences, North Carolina State University, Raleigh USA
ORCID iD: 0000-0002-4341-9675

Jose de Jesus Sánchez-González

Centro Universitario de Ciencias Biológicas y Agropecuarias, Universidad de Guadalajara, Zapopan, México

Arun Seetharam

Department of Ecology, Evolution, and Organismal Biology; Genome Informatics Facility, Iowa State University, Ames, USA
ORCID iD: 0000-0002-6789-9298

Maud Tenaillon

Génétique Quantitative et Evolution - Le Moulon, Université Paris-Saclay, INRAE, CNRS, AgroParisTech, Gif-sur-Yvette, France

Jeffrey Ross-Ibarra

Dept. of Evolution and Ecology, University of California, Davis, Davis, USA, Center for Population Biology, University of California, Davis, Davis, USA, Genome Center, University of California, Davis, Davis, USA
ORCID iD: 0000-0003-1656-4954

For correspondence: rossibarra@ucdavis.edu

Editors

Reviewing Editor
Detlef Weigel
Max Planck Institute for Biology Tübingen, Tübingen, Germany

Senior Editor
Detlef Weigel
Max Planck Institute for Biology Tübingen, Tübingen, Germany

Reviewer #1 (Public review):

Summary:

This paper examines patterns of diversity and divergence in two closely related sub-species of Zea mays. While the data are interesting and the authors have tried to exclude multiple confounding factors, many patterns cannot clearly be ascribed to one cause or another.

Strengths:

The paper presents interesting data from sets of sympatric populations of the two sub-species, maize and teosinte. This sampling offers unique insights into the diversity and divergence between the two, as well as the geographic structure of each. Many analyses and simulations to check analyses have been carried out.

Weaknesses:

The strength of conclusions that can be drawn from the analyses was low, partly because there are many strange patterns. The authors have done a good job of adding caveats, but clearly, these species do not meet many assumptions of our methods

https://doi.org/10.7554/eLife.92405.2.sa1

Author response:

The following is the authors' response to the current reviews.

> ***Public Reviews:***
>
> ***Reviewer #1 (Public review):***
>
> *Summary:*
>
> *This paper examines patterns of diversity and divergence in two closely related sub-species of Zea mays. While the data are interesting and the authors have tried to exclude multiple confounding factors, many patterns cannot clearly be ascribed to one cause or another.*
>
> *Strengths:*
>
> *The paper presents interesting data from sets of sympatric populations of the two sub-species, maize and teosinte. This sampling offers unique insights into the diversity and divergence between the two, as well as the geographic structure of each. Many analyses and simulations to check analyses have been carried out.*
>
> *Weaknesses:*
>
> *The strength of conclusions that can be drawn from the analyses was low, partly because there are many strange patterns. The authors have done a good job of adding caveats, but clearly, these species do not meet many assumptions of our methods.*

Thank you for the comments. We appreciate the multiple rounds of revision the manuscript has undergone and the work has improved as a consequence. Overall we disagree that the

patterns are strange, and have made considerable efforts to explain in the text and in our responses why the patterns make sense based on what we know about the history of Zeamays from previous research. We agree that currently available methods are not capable of answering all questions we propose adequately. This reflects both limitations with the available data for these populations (i.e. phenotypes and spatially explicit sampling), and limitations in available methods tailored to the questions at hand (spatially explicit inference of the range over which an allele is adaptive). We have made considerable effort to point out the places where our inferences are likely to have low accuracy or limited resolution. These limitations are in many ways inherent to all inferential based science and should not be considered a weak point specific to this work, nor do they take away from the fundamental conclusions, which have changed quantitatively but not qualitatively over the course of peer review.

> ***Recommendations for the authors:***
>
> *Reviewer #1 (Recommendations for the authors):*
>
> *-The manuscript should say something about the fact that range-wide PSMC does not show a decline.*

We did not use PSMC methods but instead mushi as described in the methods. On line 356 we described how the lower sample size and strong regularization are the most likely explanations for the lack of a population size decline in the rangewide samples.

> *- The manuscript should explain how rdmc was run and what "overlapping" means.*

We described how sweep intervals were inferred starting on line 823 (Methods subsection “Identifying Selective Sweeps”). Sweep regions were defined as the outermost coordinates from all populations that shared any overlap in their respectively defined sweep intervals. The details of how we ran rdmc, including all of the parameters, is described starting on line 895 (methods subsection “Inferring modes of convergent adaptation”).

> *- Figure 4: "Negative log10" is messed up*

Thank you. This has been fixed for the Version Of Record.

> *- Line 318: "accruacy"*

Thank you. We have edited this typo for the Version Of Record.

> *- New Table S3: why don't the proportions add to 1?*

These values represent what proportion of fixed differences at 0 fold sites are unique to each population. The denominator is the total number of fixed differences for each population separately, so each proportion is distinct for each population and thus should not sum to one across them. The table caption has been reworded in efforts to clarify for the Version Of Record.

The following is the authors’ response to the original reviews.

> ***Public Reviews:***
>
> ***Reviewer #1 (Public Review):***
>
> *Summary:*

> *This paper examines patterns of diversity and divergence in two closely related sub-species of Zea mays. While the patterns are interesting, the strength of evidence in support of the conclusions drawn from these patterns is weak overall. Most of the main conclusions are not supported by convincing analyses.*
>
> *Strengths:*
>
> *The paper presents interesting data from sets of sympatric populations of the two sub-species, maize and teosinte. This sampling offers unique insights into the diversity and divergence between the two, as well as the geographic structure of each.*
>
> *Weaknesses:*
>
> *There were issues with many parts of the paper, especially with the strength of conclusions that can be drawn from the analyses. I list the major issues in the order in which they appear in the paper.*
>
> *(1) Gene flow and demography.*
>
> *The f4 tests of introgression (Figure 1E) are not independent of one another. So how should we interpret these: as gene flow everywhere, or just one event in an ancestral population? More importantly, almost all the significant points involve one population (Crucero Lagunitas), which suggests that the results do not simply represent gene flow between the sub-species. There was also no signal of increased migration between sympatric pairs of populations. Overall, the evidence for gene flow presented here is not convincing. Can some kind of supporting evidence be presented?*

We agree that the standard approach to f4 tests that we employed here is not without limitations, namely, that the tests are conducted independently, while the true evolutionary history is not. While a joint demographic inference across all populations would be useful, it did not seem tractable to perform over all of our populations with currently available methods, given the number of populations being analyzed, nor does it directly address the question of interest. Our purpose for including the f4 was testing if there was more gene flow between sympatric pairs than in other comparisons (we have made that point more clear in the text near line 174. As described in the text, the distribution of Z scores is generated by pairing focal populations with all other non-focal populations across both subspecies, which means the gene flow signal of interest is marginalized over the effects of gene flow in the other non-focal populations. This is not nearly as rich as inferring the full history, but it gives us some sense of the average amount of gene flow experienced between populations and allows us to address one of our primary questions of interest when conceiving this paper — do sympatric pairs show more geneflow than other pairs? We agree with the reviewer that that answer is largely no, and the writing reflects this.

Overall, we think both points mentioned by the reviewer here; finding that most but not all tests involved Crucero Lagunitas maize, and that sympatric pairs don't show higher gene flow; nicely contributes to the overall theme in the paper — the history of both subspecies is idiosyncratic and impacted by humans in ways that do not reflect geographic proximity that we did not anticipate (see expectations near line 110). We have emphasized the connection between f4 tests and the revised rdmc results near line 653.

> *The paper also estimates demographic histories (changes in effective population sizes) for each population, and each sub-species together. The text (lines 191-194) says that "all histories estimated a bottleneck that started approximately 10 thousand generations ago" but I do not see this. Figure 2C (not 2E, as cited in the text) shows that teosinte had declines in all populations 10,000 generations ago, but some of these declines were very minimal. Maize has a similar pattern that started more recently, but the overall species*

> *history shows no change in effective size at all. There's not a lot of signal in these figures overall.*
>
> *I am also curious: how does the demographic model inferred by mushi address inbreeding and homozygosity by descent (lines 197-202)? In other words, why does a change in Ne necessarily affect inbreeding, especially when all effective population sizes are above 10,000?*

All maize populations show a decline beginning 10,000 generations ago. The smallest decline for maize is from 100,000 to 30,000. All teosinte populations show a reduction in population size. The smallest of these drops more than 70% from around 300,000 to 100,000. Three of the teosinte populations showed a reduction in population size from ~10^5 to ~10^3, which is well below 10,000. Thus all populations show declines.

These large reductions should lead to inbreeding and increased homozygosity by descent. Mushi does not specifically model these features of the data, yet as we show, simulations under the model estimated by Mushi matched the true HBD levels fairly well (Figure 2D).

The rangewide sample does not show declines, likely because there is enough isolation between populations that the reduction in variation at any given locus is not shared, and is maintained in the populations that did not experience the population decline.

> *(2) Proportion of adaptive mutations.*
>
> *The paper estimates alpha, the proportion of nonsynonymous substitutions fixed by positive selection, using two different sampling schemes for polymorphism. One uses range-wide polymorphism data and one uses each of the single populations. Because the estimates using these two approaches are similar, the authors conclude that there is little local adaptation. However, this conclusion is not justified.*
>
> *There is little information as to how the McDonald-Kreitman test is carried out, but it appears that polymorphism within either teosinte or maize (using either sampling scheme) is compared to fixed differences with an outgroup. These species might be Z. luxurians or Z. diploperennis, as both are mentioned as outgroups. Regardless of which is used, this sampling means that almost all the fixed differences in the MK test will be along the ancestral branch leading to the ancestor of maize or teosinte, and on the branch leading to the outgroup. Therefore, it should not be surprising that alpha does not change based on the sampling scheme, as this should barely change the number of fixed differences (no numbers are reported).*
>
> *The lack of differences in results has little to do with range-wide vs restricted adaptation, and much more to do with how MK tests are constructed. Should we expect an excess of fixed amino acid differences on very short internal branches of each sub-species tree? It makes sense that there is more variation in alpha in teosinte than maize, as these branches are longer, but they all seem quite short (it is hard to know precisely, as no Fst values or similar are reported).*

The section “Genetic Diversity” in the methods provides details about how luxurians and diploperennis were used as outgroups. The section “Estimating the Rate of Positive Selection, α”, in the methods includes the definition of α and full joint non-linear regression equation and the software used to estimate it (brms), and the relevant citations crediting the authors of the original method. However, some of the relevant information about the SFS construction is provided in the previous section entitled, “Genetic Diversity”. We added reference to this in results near line 800.

While we appreciate the concern that “almost all the fixed differences in the MK test will be along the ancestral branch leading to the ancestor of maize or teosinte”, this is only a problem if there aren’t enough fixed differences that are unshared between populations. This is more of a concern for maize than teosinte, which we make clear as a caveat in the manuscript in several places already. The fact that there is variation in alpha among teosinte populations is evidence that these counts do differ among pops. As we can see in the population trees in Figure 1, there is a considerable amount of terminal branch length for all the populations. Indeed if we look at the number of fixed differences at 0 fold sites across populations:

The variation in the number of fixed differences, particularly across teosinte means that a large number cannot be shared between populations. We can estimate the fixed differences unique to each subpopulation (and total count) demonstrating that, in general, there are a large number of substitutions unique to each population. This is good evidence the rangewide estimates do not reflect a lack of variation within populations, at least not for teosinte. This is now included in the supplement (Table S3).

Finally, we note that the branches leading to outgroups are likely not substantially longer than those among populations. Given our estimates of Ne, the coalescent within maize and teosinte should be relatively deep (with Ne of 30K it should be ~120K years). The divergence time between Zea mays and these outgroup taxa has been estimated at ~150K years (Chen et al. 2022). This is now mentioned in the text on line 407.

We have added a caveat about the reviewers concern for the non-independence of fixed difference for maize near line 386.

> *(3) Shared and private sweeps.*
>
> *In order to make biological inferences from the number of shared and private sweeps, there are a number of issues that must be addressed.*
>
> *One issue is false negatives and false positives. If sweeps occur but are missed, then they will appear to be less shared than they really are. Table S3 reports very high false negative rates across much of the parameter space considered, but is not mentioned in the main text. How can we make strong conclusions about the scale of local adaptation given this? Conversely, while there is information about the false positive rate provided, this information doesn't tell us whether it's higher for population-specific events. It certainly seems likely that it would be. In either case, we should be cautious saying that some sweeps are "locally restricted" if they can be missed more than 85% of the time in a second population or falsely identified more than 25% of the time in a single population.*

The reviewer brings up a worthwhile point. The simulation results indeed call into question how many of the sweeps we claim are exclusive to one population actually are. This caveat is already made, but we now make clearer the reviewer’s concern regarding the high false negative rate (near line 299). However, if anything this suggests sweeps are shared even more often than what is reported. One of the major takeaways from the paper is that convergent adaptation is more common than we expected. The most interesting part about the unique sweeps is the comparison between maize and teosinte. While the true proportions may vary, the relatively higher proportion of sweeps exclusive to one population in teosinte compared to maize is unlikely to be affected by false negatives, since the accuracy to identify sweeps pretty similar across subspecies (though perhaps with some exceptions for the populations with stronger bottlenecks). Further, these criticisms are specific to the raisd results. All sweeps shared across multiple populations were analyzed using rdmc. After adjustments made to the number of proposed sites for selection (see response below), there is good agreement between the raisd and rdmc results — the regions we proposed as selective

sweeps with raisd all show evidence convergence using rdmc. Recall too that rdmc uses a quite different approach to inference — all populations are used jointly, labelling those that did and did not experience the sweep. If sweeps were present in populations that were labeled as neutral (or vice versa), this would weaken the power to infer selection at the locus. Much of the parameter space we explored is for quite weak selection, and the simulated analysis shows we are likely to miss those instances, often entirely. For strong sweeps, however, our simulations show we have appreciable accuracy.

Together, there is reason to be optimistic about our detection of strong shared sweeps and that the main conclusions we make are sound.

Finally, we note that we are unaware of any other empirical study that has performed similar estimates of the accuracy of the sweep calling in their data (as opposed to using simulations). We thus see these analyses as a significant contribution towards transparency that is completely lacking from most papers.

> *A second, opposite, issue is shared ancestral events. Maize populations are much more closely related than teosinte (Figure 2B). Because of this, a single, completed sweep in the ancestor of all populations could much more readily show a signal in multiple descendant populations. This is consistent with the data showing more shared events (and possibly more events overall). There also appear to be some very closely (phylogenetically) related teosinte populations. What if there's selection in their shared ancestor? For instance, Los Guajes and Palmar Chico are the two most closely related populations of teosinte and have the fewest unique sweeps (Figure 4B). How do these kinds of ancestrally shared selective events fit into the framework here?*

The reviewer brings up another interesting point and one that likely impacts some of our results.

As the reviewer describes, this is an issue that is of more concern to the more closely related populations and is less likely to explain results across the subspecies. We have added this as a caveat (near line 456). As is clear in the writing, sharing across subspecies is our primary interest for the rdmc results.

> *These analyses of shared sweeps are followed by an analysis of sweeps shared by sympatric pairs of teosinte and maize. Because there are not more events shared by these pairs than expected, the paper concludes that geography and local environment are not important. But wouldn't it be better to test for shared sweeps according to the geographic proximity of populations of the same sub-species? A comparison of the two sub-species does not directly address the scale of adaptation of one organism to its environment, and therefore it is hard to know what to conclude from this analysis.*

We did not intend to conclude that local adaptation is not important. Especially for teosinte, we report and interpret evidence that many sweeps are happening exclusively to one population, which is consistent with the action of location adaptation and consistent with some of our expectations.

More directly, this is another instance of us having clear hypotheses going into the paper and constructing specific analyses to test them. As we explain in the paper, we expected the scale of local adaptation to be very small, such that subspecies growing next to each other have more opportunities to exchange alleles that are locally adapted to their shared environment. The analysis we conducted makes sense in light of this expectation. We considered conducting tests regarding geographic proximity, but there is limited power with the number of populations we have within subspecies, and the meaning of the tests is unclear if all populations of both subspecies are naively included together. This analysis shows that, at

least for sweeps and fixations, adaptation is larger than a single location. While it may not be a complete description on its own, the work here does provide information about the scale of adaptation and is useful to our overall claims and objectives of the paper. As mentioned in the paper, the story might be very different if we were to study through a lens of polygenic adaptation. We also now include in the discussion in several places mention of where broader sampling could improve inference.

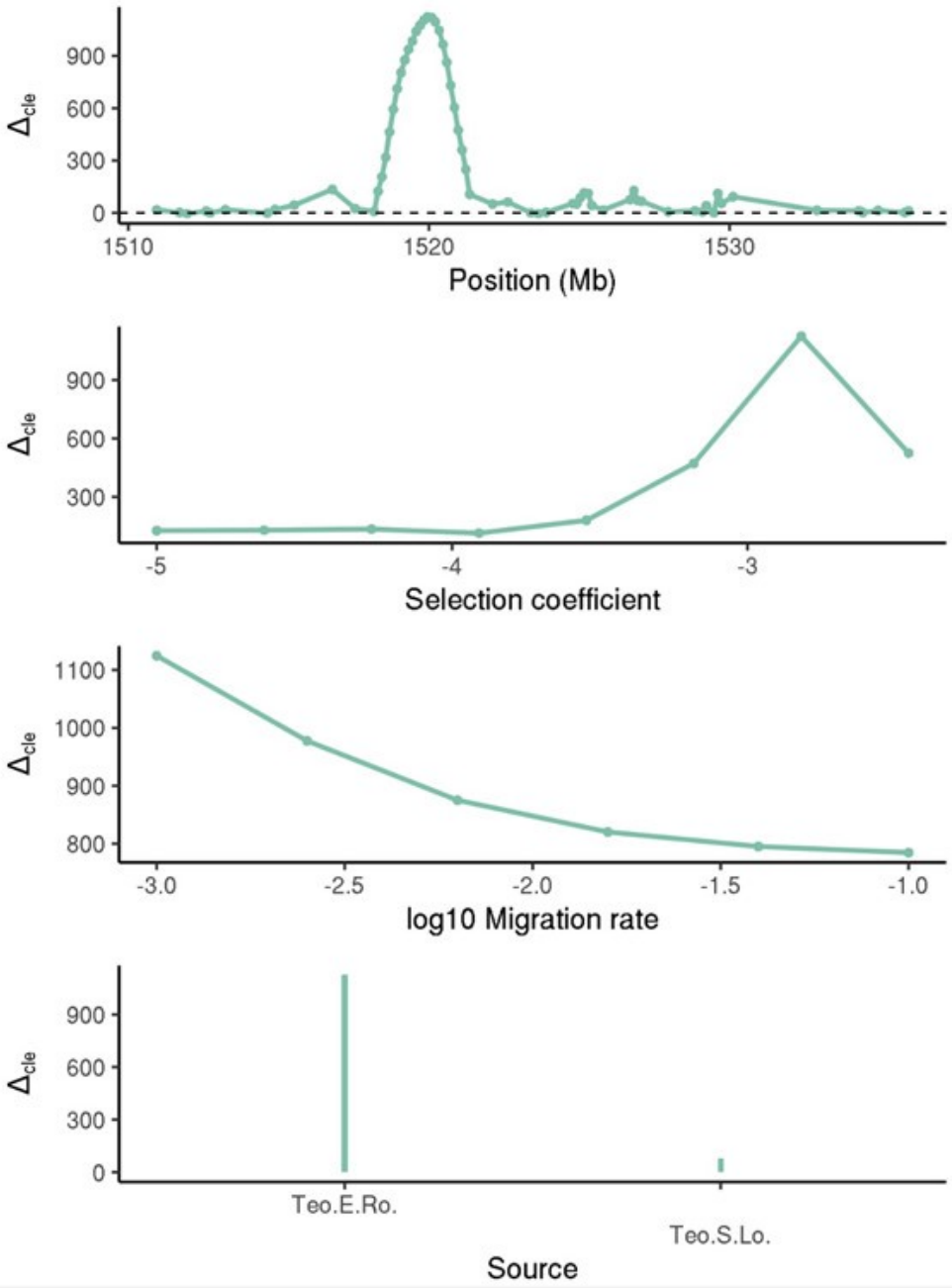
> *(4) Convergent adaptation*
>
> *My biggest concern involves the apparent main conclusion of the paper about the sources of "convergent adaptations". I believe the authors are misapplying the method of Lee and Coop (2017), and have not seriously considered the confounding factors of this method as applied. I am unconvinced by the conclusions that are made from these analyses.*
>
> *The method of Lee and Coop (referred to as rdmc) is intended to be applied to a single locus (or very tightly linked loci) that shows adaptation to the same environmental factor in different populations. From their paper: "Geographically separated populations can convergently adapt to the same selection pressure. Convergent evolution at the level of a gene may arise via three distinct modes." However, in the current paper, we are not considering such a restricted case. Instead, genome-wide scans for sweep regions have been made, without regard to similar selection pressures or to whether events are occurring in the same gene. Instead, the method is applied to large genomic regions not associated with known phenotypes or selective pressures.*
>
> *I think the larger worry here is whether we are truly considering the "same gene" in these analyses. The methods applied here attempt to find shared sweep regions, not shared genes (or mutations). Even then, there are no details that I could find as to what constitutes a shared sweep. The only relevant text (lines 802-803) describes how a single region is called: "We merged outlier regions within 50,000 Kb of one another and treated as a single sweep region." (It probably doesn't mean "50,000 kb", which would be 50 million bases.) However, no information is given about how to identify overlap between populations or sub-species, nor how likely it is that the shared target of selection would be included in anything identified as a shared sweep. Is there a way to gauge whether we are truly identifying the same target of selection in two populations?*
>
> *The question then is, what does rdmc conclude if we are simply looking at a region that happened to be a sweep in two populations, but was not due to shared selection or similar genes? There is little testing of this application here, especially its accuracy. Testing in Lee and Coop (2017) is all carried out assuming the location of the selected site is known, and even then there is quite a lot of difficulty distinguishing among several of the non-neutral models. This was especially true when standing variation was only polymorphic for a short time, as is estimated here for many cases, and would be confused for migration (see Lee and Coop 2017). Furthermore, the model of Lee and Coop (2017) does not seem to consider a completed ancestral sweep that has signals that persist into current populations (see point 3 above). How would rdmc interpret such a scenario?*
>
> *Overall, there simply doesn't seem to be enough testing of this method, nor are many caveats raised in relation to the strange distributions of standing variation times (bimodal) or migration rates (opposite between maize and teosinte). It is not clear what inferences can be made with confidence, and certainly the Discussion (and Abstract) makes conclusions about the spread of beneficial alleles via introgression that seem to outstrip the results.*

We have fixed the “50,000 Kb” typo.

There are several important points the reviewer makes here worth considering. First and most importantly, the method of Lee and Coop (2017) actually does include sites as part of the composite likelihood calculation. For computational feasibility, the number of positions we initially considered was 20 (20 different positions along the input sequence were proposed as the site of the shared beneficial mutation). In efforts to further address the reviewer's concern about adaptive mutations at distinct loci, we have increased the number of proposed selected sites to 200. This fact should greatly diminish the reviewer's concern that we are picking up independent sweeps that happened at different nucleotide positions in the same region — evidence for a beneficial mutation must be shared by the selected populations at a proposed site. As the revisions show, this has modified the results of our paper in a number of ways, including changing all of the previous neutral regions to shared via standing variation or migration. Despite these changes, our previous conclusions are intact, including the pattern that migration rates are high when maize populations share the sweep. Relatedly, we disagree with the reviewer's characterization of the migration results. The pattern is quite clear and makes sense — when a maize population is involved in the sweep, migration rate is inferred to be high. Sweeps exclusive to teosinte are rarer and are inferred to have a low migration rate. This relates directly to the idea that humans have moved maize relatively rapidly across the landscape.

We have now included a plot showing how the difference between the maximum composite likelihood (CLE) site compares to the next highest CLE site varies across our inferences (Figure S8), which strongly suggests that patterns are not muddled across multiple loci, but are centered at a focal region where the beneficial allele is inferred to be located. While there are too many to show in the manuscript across all sweeps, here is a nice example of what inference looks like for one of the proposed sweep regions.

Author response image 1.



Furthermore, the situation the reviewer is describing would be selection acting on independent mutations (mutations at different loci), which would not create an increase in the amount of allele frequency covariance above and beyond what would be expected by drift under the migration and standing variation models.

We also note that we are not alone in applying this approach to shared outlier signals in the absence of known genes; indeed the authors of the DMC method have applied it to regions of shared outlier signal themselves (e.g. https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1008593).

> ***Reviewer #2 (Public Review):***
>
> *Summary:*
>
> *The authors sampled multiple populations of maize and teosinte across Mexico, aiming to characterise the geographic scale of local adaptation, patterns of selective sweeps, and modes of convergent evolution between populations and subspecies.*
>
> *Strengths & Weaknesses:*
>
> *The population genomic methods are standard and appropriate, including Fst, Tajima's D, α, and selective sweep scans. The whole genome sequencing data seems high quality. However, limitations exist regarding limited sampling, potential high false-positive sweep detection rates, and weak evidence for some conclusions, like the role of migration in teosinte adaptation.*
>
> *Aims & Conclusions:*
>
> *The results are interesting in supporting local adaptation at intermediate geographic scales, widespread convergence between populations, and standing variation/gene flow facilitating adaptation. However, more rigorous assessments of method performance would strengthen confidence. Connecting genetic patterns to phenotypic differences would also help validate associations with local adaptation.*
>
> *Impact & Utility:*
>
> *This work provides some of the first genomic insights into local adaptation and convergence in maize and teosinte. However, the limited sampling and need for better method validation currently temper the utility and impact. Broader sampling and connecting results to phenotypes would make this a more impactful study and valuable resource. The population genomic data itself provides a helpful resource for the community.*
>
> *Additional Context:*
>
> *Previous work has found population structure and phenotypic differences consistent with local adaptation in maize and teosinte. However, genomic insights have been lacking. This paper takes initial steps to characterise genomic patterns but is limited by sampling and validation. Additional work building on this foundation could contribute to understanding local adaptation in these agriculturally vital species.*

We appreciate the reviewer's thoughtful reading of the paper and scrutiny. We hope that the added caveats made in response to reviewer 1 (as well as the previous rounds of peer review) will provide readers with the proper amount of skepticism in the accuracy of some of our initial sweep results, while also demonstrating that many of our conclusions are robust to the concerns raised over the various stages of review.

We agree with the reviewer that better sampling and the incorporation inference about phenotypic data would be excellent additions, but the information is not available for the studied populations, and is outside scope of this paper.

> ***Recommendations for the authors:***
>
> ***Reviewer #1 (Recommendations For The Authors):***
>
> *- Sometimes alpha is described as a rate, and sometimes as a proportion. The latter is correct.*

We have updated this. Thanks.

> *- Line 79: are they really "discrete" populations?*

The teosinte populations sampled are all clearly separated from each other and are physically discrete. The maize population samples came from individual farmer fields. Traditional maize is grown as open-pollinated (outcrossing) populations, and farmers save seed for subsequent generations. An individual farmer's field thus behaves as a discrete population for our purposes, impacted of course by gene flow, selection, and other evolutionary processes.

> *- Lines 418-420: "Large genomes may lead to more soft sweeps, where no single mutation driving adaptive evolution would fix (Mei et al. 2018)." I'm not sure I understand this statement. Why is this a property of genome size?*

Mei et al. 2018 lay out the logic, but essentially they present data arguing that the total number of functionally relevant base pairs increases with genome size (less than linearly). If true, genomes with a large number of potentially functional bp are more likely to undergo soft sweeps (see theory by Hermisson and Pennings cited in Mei et al. 2018).

> *- Lines 500-1: selection does not cause one to underestimate effective population sizes. Selection directly affects Ne. I'm not sure what biases the sentences on lines 502-508 are trying to explain.*

We have simplified this section. Not accounting for linked selection (especially positive selection) results in a biased inference of demographic history. See Marsh and Johri (2024) for another example. https://doi.org/10.1093/molbev/msae118

> *- Line 511-3: does Uricchio et al. (2019) show any difference in the estimate of alpha from Messer and Petrov (2013) when taking background selection into account?*

What we initially wrote was incorrect. The aMK method of Messer and Petrov (2013) accounts for weakly deleterious polymorphisms, but it does not account for positively selected ones. We have updated this text and suggested our method may underestimate alpha if positively selected segregating alleles are common (near line 539).

> *- Lines 598-599: "which would limit the rate of new and beneficial mutations." I don't understand this - shouldn't a bottleneck only affect standing variation? Why would a bottleneck affect new mutations?*

This is simply to say that during the low Ne period of a bottleneck, fewer total mutations (and therefore beneficial mutations) will be generated since there are fewer individuals for mutations to occur in. We have changed "rate" to amount to clarify we do not mean the mutation rate itself.

> ***Reviewer #2 (Recommendations For The Authors):***
>
> *Experiments/Analyses:*
>
> *(1) Consider simulating polygenic adaptation in addition to hard and soft sweeps to see if this improves the power to detect adaptive signatures shared between populations. This could involve simulating the coordinated change in allele frequencies across many loci to match a specified shift in trait value due to selection. The ability to detect shared polygenic adaptation between population replicates could be assessed using methods*

> *tailored to polygenic signals, such as the Polygenic Selection Score approach. Comparing the power to detect shared polygenic adaptation versus shared hard and soft sweeps would provide further insight into what adaptive modes current methods can uncover. If the power to detect shared polygenic adaptation is very low, the extent of shared adaptation between populations may be even more common than currently inferred. Adding simulations of polygenic adaptation would strengthen the study.*

While this would be a worthwhile undertaking in general, it would be a considerable amount of work outside of the scope and aims of this paper.

> *(2) Explore using machine learning approaches like S/HIC to improve power over summary statistic methods potentially.*

We in fact put considerable effort into applying diplo S/HIC before switching to raisd for this project. While predictions on simulations had good power to detect sweeps, we found that applying to our actual data had a dubious number of windows classified as sweeps (e.g. >90% of the genome), which we believed to be false positives. We speculated that this may have to do with sensitivity to demographic or other types of misspecification in the simulations, such as our choice of window sizes compared to local recombination rates. It would likely be fruitful to our further efforts into using machine learning methods for maize and teosinte, but a deeper exploration of the right hyper parameters and simulation choices is likely needed to apply them effectively.

> *(3) Increase geographic sampling density, if possible, especially near population pairs showing high differentiation, to better understand the scale of local adaptation.*

We agree this would be valuable research. Hopefully this work inspires further efforts into the question of the spatial and temporal scales of local adaptation with more ambitious spatial sampling designed at the onset

> *Writing/Presentation:*
>
> *(1) Provide more intuition about the biological interpretation of the migration rates inferred under the migration model of convergence. What do the rates imply about the amount or timing of gene flow?*

We have expanded the discussion sections (starting near line 653) to elaborate on the migration results and connect the rdmc and f4 tests more explicitly. The timing of gene flow is more challenging to address directly with the approaches we used, but we agree it would be interesting to explore more in future papers.

> *(2a) Expand the discussion of power limitations and the need for simulation tests. Consider adding ROC curves for sweep detection on simulated data. The relatively low proportion of shared selective sweeps between population replicates highlights limitations in the power to detect sweeps, especially incomplete or soft sweeps. I think it would be a good idea to expand the discussion of the power tradeoffs shown in the simulation analyses. In particular, the ROC curves in Figure S4 clearly show how power declines for weaker selection coefficients across the different sweep types. I suggest making these ROC curves part of the main figures to feature the issue of power limitations more prominently.*
>
> *(2b) The discussion would benefit from commenting on how power changes across the sweep simulation scenarios. Adding a summary figure to visualise the effects of sweep type, selection strength, and frequency on detectability could further clarify the power constraints. Stating the proportion of sweeps likely missed strengthens the argument*

> *that sharing adaptive alleles is likely even more common than inferred. Discussing power will also motivate the need for developing methods with improved abilities to uncover incomplete and soft sweeps.*

While these are useful suggestions (2a and 2b), the aim of this paper at its core is empirical, and was not intended to give an exhaustive analysis of the power to detect sweeps. We report what parts of the analysis may be impacted by low power and what aspects of our inferences have higher uncertainty due to power. We agree that there is more work to be done to improve methods to detect selection given our findings (see below concerning our efforts to use machine learning as well). While we do not highlight this in the paper, we also note that ours is one of extremely few empirical studies that actually perform power analyses on real data (as opposed to simulations). We think this extra transparency by itself is of substantial utility to the community in demonstrating that the results from simulation studies performed in publications describing a method do not necessarily translate well to empirical data.

> *(3) Improve clarity in describing f4 test results. Consider visualising results on a map to show spatial patterns.*

We have expanded the discussion concerning f4 tests (see several comments to reviewer 1). We are not clear on how to effectively visualize f4 spatially, but hope the updates have made the results more clear.

> *Minor:*
>
> *- Increase the font size of figure axis labels for improved readability.*

We have looked over and figures and increased font sizes where possible.

> *- Add units to selection coefficient axis labels in Figure 5.*

Selection coefficients are derived in Lee and Coop (2017) from classical population genetics theory. They do not have units, but denote the relative fitness advantage of the heterozygous genotype carrying the beneficial mutation of interest.

> *- Fix the typo 'cophenetic' in Figure S3 caption.*

Fixed. Thank you.

https://doi.org/10.7554/eLife.92405.2.sa0