Selective sorting of ancestral introgression in maize and teosinte along an elevational cline

Erin Calfee[1,2*], Daniel Gates[2], Anne Lorant[3,5], M. Taylor Perkins[2,6], Graham Coop[1,2,†], and Jeffrey Ross-Ibarra[1,2,4,†]

1 Center for Population Biology, University of California, Davis, California, USA
2 Department of Evolution and Ecology, University of California, Davis, California, USA
3 Department of Plant Sciences, University of California, Davis, California, USA
4 Genome Center, University of California, Davis, California, USA
5 Laboratoire de Biologie Moléculaire et Cellulaire du Cancer, Hôpital Kirchberg, Luxembourg (current address)
6 Plant Molecular and Cellular Biology Program, University of Florida, Gainesville, Florida, USA (current address)
† co-mentors.

*** ecalfee@ucdavis.edu**

Abstract

While often deleterious, hybridization can also be a key source of genetic variation and pre-adapted haplotypes, enabling rapid evolution and niche expansion. Here we evaluate these opposing selection forces on introgressed ancestry between maize (*Zea mays* ssp. *mays*), and its wild teosinte relative. Introgression from ecologically diverse teosinte may have facilitated maize's global range expansion, in particular to challenging high elevation regions ($>$ 1500 m). We generated low-coverage genome sequencing data for 348 maize and *mexicana* individuals to evaluate patterns of introgression in 14 sympatric population pairs, spanning the elevational range of *Zea mays* ssp. *mexicana*, a teosinte endemic to the mountains of Mexico. While recent hybrids are commonly observed in sympatric populations and *mexicana* demonstrates fine-scale local adaptation, we find that the majority of *mexicana* ancestry tracts introgressed >1000 generations ago. This *mexicana* ancestry seems to have maintained much of its diversity and likely came from a common ancestral source, rather than contemporary sympatric populations, resulting in relatively low F_{ST} between *mexicana* ancestry tracts sampled from geographically distant maize populations. Introgressed *mexicana* ancestry is reduced in lower-recombination rate quintiles of the genome and around domestication genes, consistent with pervasive selection against introgression. However, we also find *mexicana* ancestry increases across the sampled elevational gradient and that high introgression peaks are most commonly shared among high-elevation maize populations, consistent with introgression from *mexicana* facilitating adaptation to the highland environment. In the other direction, we find patterns consistent with adaptive and clinal introgression of maize ancestry into sympatric *mexicana* at many loci across the genome, suggesting that maize also contributes to adaptation in *mexicana*, especially at the lower end of its elevational range. In sympatric maize, in addition to high introgression regions we find many genomic regions where selection for local adaptation maintains

steep gradients in introgressed *mexicana* ancestry across elevation, including at least two inversions: the well-characterized *Inv4m* and a new 3 Mb inversion *Inv9f* surrounding the *macrohairless1* locus on chromosome 9. The bulk of our ancestry selection outliers show no signals of sweeps or local sourcing from sympatric populations and so likely represent ancestral introgression sorted by selection, resulting in correlated but distinct outcomes of introgression in different contemporary maize landrace populations.

Introduction

Interbreeding between partially diverged species or subspecies can result in admixed individuals with low fitness, e.g. due to hybrid incompatibilities [76, 77, 115]. Consistent with the view that hybridization is often deleterious, a growing number of species show evidence of pervasive selection against introgressed ancestry [1, 10, 26, 39, 45, 46, 59, 65, 87, 89]. At the same time, introgression can be a source of novel genetic variation and efficiently introduce haplotypes carrying sets of locally-adapted alleles, with the potential for rapid adaptation to new ecological challenges [71]. Indeed, admixture has been linked to adaptive species radiations and/or rapid niche expansions in a number of natural systems, including mosquitoes [28], *Drosophila* [97], butterflies [26], cichlids [58], sunflowers [82], wild tomatoes [68] and yeast [25, 100]. In addition, introgression from wild relatives has facilitated the broad range expansions of multiple domesticated crops (reviewed in [78] and [43]), and gene flow from crops back into their wild relatives has in some cases opened up novel 'weedy' niches [27].

Maize (*Zea mays* ssp. *mays*) is an ideal system to study selection on admixed ancestry and the effects on range expansion, as it has colonized nearly every human-inhabited ecosystem around the world [38] and interbreeds with a number of wild relatives genetically adapted to distinct ecologies [41, 57]. In Mexico, highland maize represents an early major niche expansion that may have been facilitated by introgression. Approximately 9 thousand years ago, maize (*Zea mays* ssp. *mays*) was domesticated in the Balsas River Valley in Mexico from a lowland-adapted sub-species of teosinte (*Zea mays* ssp. *parviglumis* [74]), which grows readily at sea level and up to about 2000 meters [35]. In contrast, *Zea mays* ssp. *mexicana*, which diverged from *parviglumis* about 60 thousand years ago [85], is endemic to highland regions in Mexico (∼1500-3000 meters in elevation) where it has adapted to a number of ecological challenges: a cooler, drier climate with higher UV intensity, different soil nutrient composition, and a shorter growing season necessitating earlier flowering times [2, 30, 31, 72, 83].

Maize was introduced as a crop to the mountains of Mexico around 6.2 thousand years ago [73], and it is thought that gene flow from *mexicana* assisted in adaptation to high elevation selection pressures. Highland maize and *mexicana* share a number of putatively adaptive phenotypes [109, 110], including earlier flowering times for the shorter growing season [83], purple anthocyanin-based pigmentation which shields DNA from UV damage [95] and increases solar heat absorption [5], and macrohairs on the leaf and stem sheath, which are thought to increase herbivore defense [64] and/or heat maintenance in colder environments [49]. Earlier studies using 50K SNP-chip data for highland populations [42] or genomewide data for a small number of individuals [34, 105], have shown that highland maize landraces have experienced significant admixture from *mexicana*, reaching high frequency at some loci, consistent with adaptive introgression.

While some highland and locally-adapted alleles may be beneficial to maize, many introgressed *mexicana* alleles, especially those affecting domestication traits, should be selected against by farmers growing maize landraces. In addition, maize alleles introgressed into *mexicana* should be selected against because maize has accumulated genetic load from reduced population sizes during domestication [105] and because

domestication traits generally reduce fitness in the wild [17, 61, 96], e.g. loss of disarticulation and effective seed dispersal [109].

In this study, we generate whole genome sequencing to investigate genomic signatures of admixture and selection in paired maize landrace and sympatric *mexicana* populations, sampled from 14 locations across an elevational gradient in Mexico. This expanded sampling of sympatric maize and *mexicana* populations across Mexico, combined with genomewide data and a well-parameterized null model, improves our ability to more formally test for adaptive introgression and identify likely source populations. The source of introgression is of interest, as teosinte demonstrates local adaptation to different niches within the highlands and there is significant genetic structure between *mexicana* ecotypes [29, 30, 67, 79, 109]. Thus we can test whether local *mexicana* populations are the ongoing source for geographically-restricted locally adaptive haplotypes. We use this comprehensive genomic dataset to characterize the bi-directional timing and origin of introgression and evaluate the patterns and scale of natural selection for and against admixture between these taxa.

Results/Discussion

Genomewide *mexicana* ancestry is structured by elevation

We sampled paired sympatric populations from 14 geographically dispersed locations to assess the extent of gene flow between maize and *mexicana* in Mexico. Maize today is grown across the entire elevational range of its wild relatives, from sea-level up to 4000 meters [94]. Our sampled sites range from 1547-2600 meters in elevation, which spans a large portion of *mexicana*'s range and exceeds the upper elevational range for maize's wild ancestor, *parviglumis* (Fig 1). For each of 14 maize/*mexicana* sympatric sample locations, we resequenced 7-15 individuals per subspecies. We additionally sequenced 43 *mexicana* individuals from 3 allopatric reference populations, totalling 348 low-coverage genomes (mean ∼1x). To this dataset, we added 55 previously published high-coverage genomes from Palmar Chico [13] as an allopatric maize reference population.

Principal components analysis of genetic diversity clearly separates maize and *mexicana*, with putative admixed individuals from sympatric populations having intermediate values along PC1 (PCAngsd, Fig S2). To estimate genomewide ancestry proportions for each individual, we ran NGSAdmix [91] with K=2 genetic clusters and genotype likelihoods for all maize and *mexicana* individuals. The two genetic clusters clearly map onto maize and *mexicana* ancestry, with no indication of gene flow into the allopatric maize reference population but small amounts of maize ancestry in two of the three allopatric *mexicana* populations (Fig 2A). Furthermore, we find a positive association between ancestry proportion and elevation (km), with higher *mexicana* ancestry at higher elevations in both sympatric maize ($\beta = 0.196, P = 1.42 \times 10^{-29}$) and sympatric *mexicana* ($\beta = 0.197, P = 4.38 \times 10^{-19}$) individuals (Fig 2B). Increasing *mexicana* ancestry at higher elevations is consistent with selection favoring *mexicana* ancestry at higher elevations, but could also be due to purely demographic processes, e.g. a higher density of (wind-dispersed) *mexicana* pollen at higher elevations, or increased gene flow from non-admixed maize populations at lower elevations.

Origin and timing of introgression

If *mexicana* ancestry found in contemporary landrace genomes facilitated maize's colonization of the highlands approximately 6.2 thousands years ago [73], we would expect introgressed ancestry tracts to be short, due to many generations of recombination, and possibly to be derived from an ancient source population common

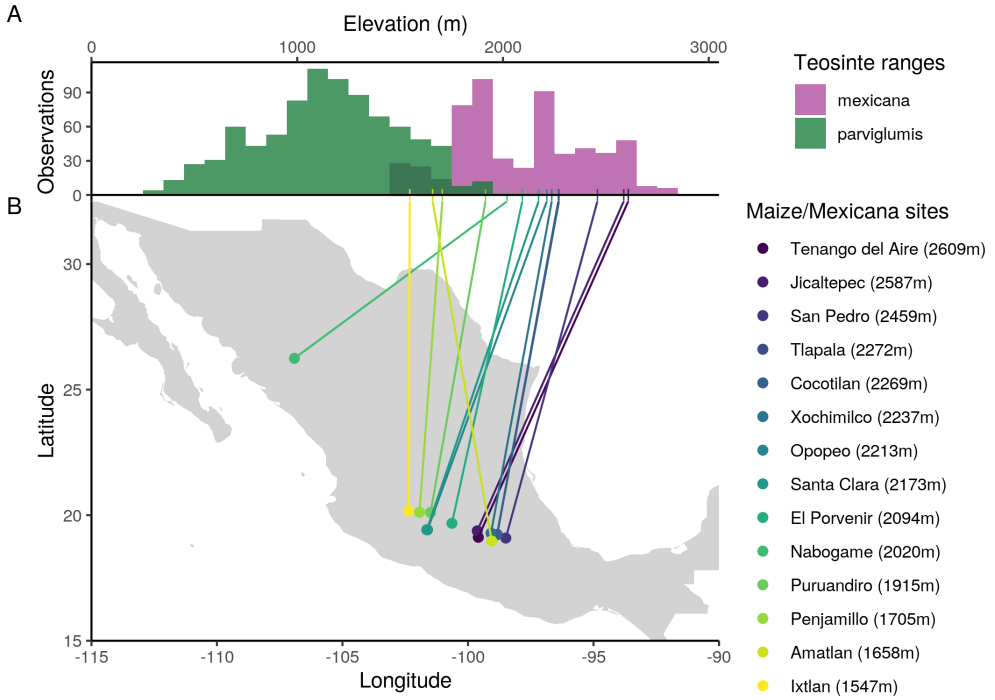

Fig 1. Sampled sympatric maize/*mexicana* populations compared to distribution of teosintes (A) Elevational range of teosintes based on historical occurrence data (1842-2016) from [35]. (B) Geographic location and elevation of contemporary sympatric maize and *mexicana* population pairs sampled across 14 sites in Mexico.

to many present-day maize populations. To test these predictions, we estimated local ancestry across the genome for individuals from each sympatric maize and *mexicana* population using a hidden Markov model based on read counts ([14] see methods). This method simultaneously estimates local ancestry and the generations since admixture for the focal population.

Admixture is generally old, with median estimates of 1203 generations for sympatric maize populations and 718 generations for sympatric *mexicana* populations (Fig S11). Because this HMM fits a single-pulse model to what was almost certainly multiple admixture events over time, we caution against over-interpretation of exact dates. Multiple pulses or ongoing gene flow biases estimates towards the more recent pulse(s) [56, 62] and even old estimates do not exclude the possibility of limited more recent admixture. These single-pulse approximations do, however, provide evidence that a large proportion of the introgression, especially into maize, is found on short ancestry tracts and therefore relatively old.

To identify likely source population(s) for introgressed ancestry, we compared F_{ST} between all sympatric populations using only reads from high-confidence homozygous ancestry tracts (posterior > 0.8) for maize and *mexicana* ancestry separately. We find that most *mexicana* ancestry in maize resembles other *mexicana* ancestry introgressed into other maize populations, rather that *mexicana* ancestry from the local sympatric *mexicana* population (Fig 3). This finding is consistent with most introgressed ancestry being drawn from a communal source population, but none of the sympatric *mexicana* populations have low enough F_{ST} to tracts introgressed into maize to be a recent source. While we cannot rule out recent introgression from an unsampled source population, the timing of our admixture estimates is more consistent with divergence of *mexicana*

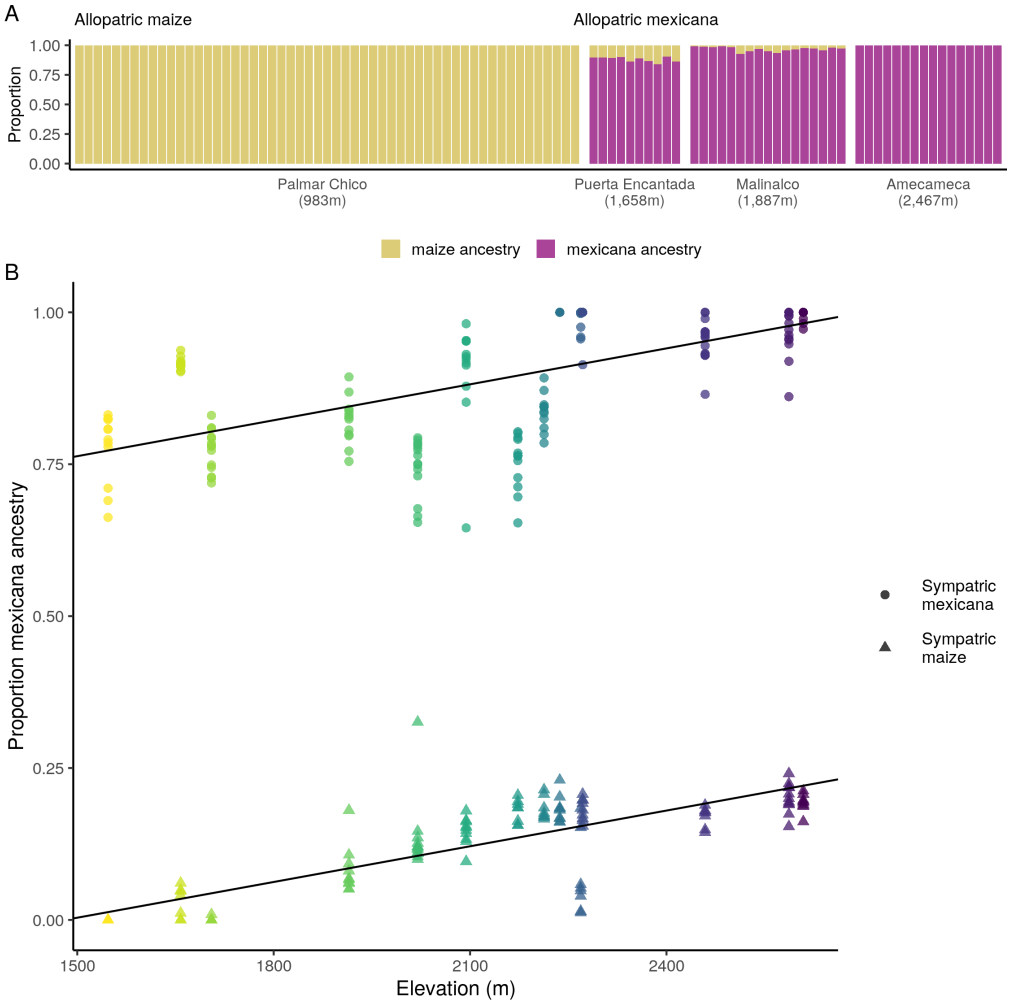

Fig 2. Distribution of *mexicana* ancestry by elevation (A) Genomewide ancestry estimates (NGSAdmix) for allopatric maize and *mexicana* reference individuals, grouped by sampling location. (B) Genomewide *mexicana* ancestry estimates (NGSAdmix) for sympatric maize and *mexicana* individuals (n = 305) along an elevational gradient, colored by sampling location. Lines show best linear model fit for *mexicana* ancestry by elevation for each subspecies separately.

ancestry, once introgressed into a maize background, from its original source population(s) (Fig S11). Additionally, *mexicana* ancestry tracts in maize have only slightly reduced genetic diversity (π, Fig S16), meaning many *mexicana* haplotypes have introgressed into maize at any given locus, with no evidence of a strong historical bottleneck.

Two lower elevation maize populations are an exception to this general pattern: Ixtlan and Penjamillo. These populations have higher F_{ST} between their introgressed ancestry tracts and other *mexicana* tracts in maize, more recent timing of admixture estimates, and reduced genetic diversity (Figs S15-S16). These patterns could be caused by small population sizes and more recent independent admixture, although F_{ST} does not identify a likely *mexicana* source population. Consistent with this interpretation, we have evidence that local maize at Ixtlan is at least partially descended from recently introduced commercial seed (relayed by local farmers [42]).

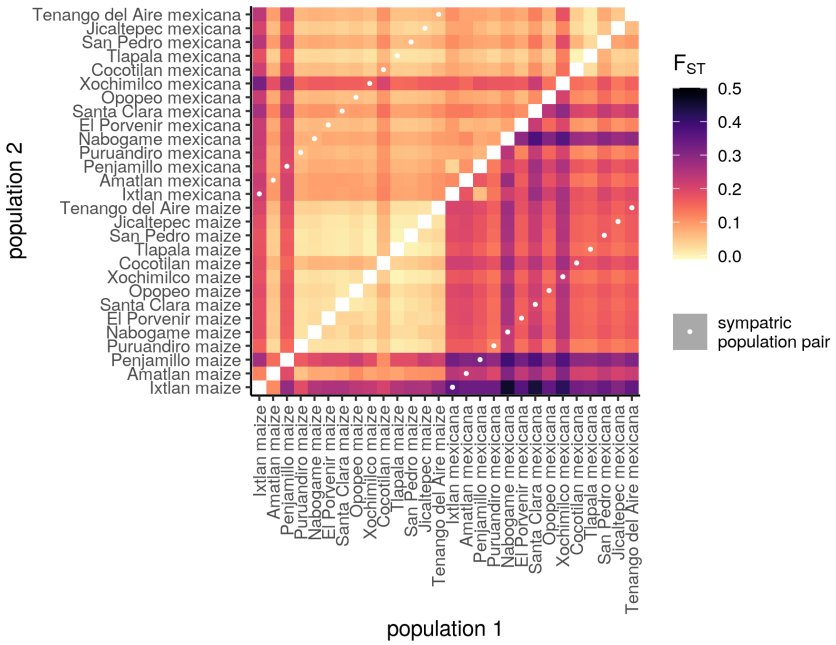

Fig 3. F_{ST} between ancestry tracts from different populations Pairwise F_{ST} between maize ancestry tracts from population 1 (x-axis) and population 2 (y-axis) are shown in the upper left triangle, while F_{ST} estimates for *mexicana* ancestry tracts are shown in the lower right triangle. Populations are sorted by subspecies, then elevation. Sympatric maize-*mexicana* population pairs are highlighted with a white dot.

The lack of a clear reduction in F_{ST} for *mexicana* ancestry tracts between sympatric population pairs, combined with older timing of admixture estimates, indicates that while contemporary hybridization may occur in the field between maize crops and adjacent *mexicana* populations, this is not the source for the bulk of the introgressed ancestry segregating in highland maize. Instead, we propose that the majority of *mexicana* ancestry in maize derives from admixture over 1000 years ago, possibly from a diverse set of *mexicana* source populations over a large geographic and temporal span, and the resulting ancestry tracts are now distributed across different contemporary maize populations. These genomewide average F_{ST} results, however, do not exclude the possibility that particular regions were introgressed from one or more distinct, possibly local, source populations.

While we also analyzed F_{ST} within high-confidence maize ancestry tracts, we found that maize ancestry is too homogeneous to make inferences about potential admixture source populations of maize into *mexicana* (Figs 3, S15).

Selection against introgression genomewide

When there is widespread selection against introgressing variants at many loci across the genome, selection will more efficiently remove linked ancestry in regions of the genome with lower recombination rates, which creates a positive relationship between local recombination rate and the proportion of introgressed ancestry [1, 10, 26, 39, 45, 46, 59, 65, 87, 89, 102]. To test whether such negative selection is shaping patterns of introgression genomewide in sympatric maize and *mexicana*, we first divided the genome into quintiles based on the local recombination rates for 1 cM windows. We then ran NGSAdmix on the SNPs within each quintile separately, using

K=2 clusters, to estimate maize and *mexicana* ancestry proportions. We used a recombination map from maize [66], which is likely to be correlated with other *Zea* subspecies at least at the level of genomic quintiles, but a limitation of this analysis is that we do not have a recombination map for hybrid populations which means that e.g. segregating structural inversions will not necessarily show low recombination rates.

Our results from sympatric maize landraces are consistent with selection against *mexicana* introgression at many loci genomewide, resulting in lower introgressed ancestry in regions of the genome with lower recombination rates (Fig 4A). We find a positive Spearman's rank correlation between recombination rate quintile and mean introgressed *mexicana* ancestry proportion ($\rho = 1,\ \mathrm{CI}_{95}[0.85,\ 1.00]$), and a similar analysis using f_4 statistics replicates this result (see methods, Fig S6-S7 and S5 Table). The higher elevation maize populations show this pattern most starkly; while all individuals have low *mexicana* ancestry for the lowest recombination rate quintile, some high elevation populations have individuals with over 40% introgressed ancestry for the highest recombination rate quintile (Fig 4B). Using a linear-model fit, we found a significant interaction between recombination rate quintile and the slope of ancestry across elevation in sympatric maize (S3 Table). This is again consistent with low-recombination rate regions having a stronger effect of linked selection reducing *mexicana* ancestry, with higher elevation maize landraces either experiencing larger amounts of gene flow or retaining more ancestry due to adaptive processes in high recombination regions (Fig S4).

Because recombination rate is positively correlated with gene density in *Zea* [88], we also tested the Spearman's rank correlation between quintiles defined by coding base pairs per cM and their proportion introgressed *mexicana* ancestry. Again we found evidence supporting pervasive selection against introgression (Fig S3, $\rho = -1,\ \mathrm{CI}_{95}[-1,\ -0.85]$).

In contrast, sympatric *mexicana* shows an unexpected negative relationship between recombination rate and introgression from maize, with more *mexicana* ancestry (lower introgression) in the highest recombination rate regions of the genome ($\rho = 1,\ \mathrm{CI}_{95}[0.9,\ 1]$). Correlations with coding bp per cM and based on f_4 statistics corroborate this pattern (see S3, S8 and S9). While one possible explanation is that introgressing maize ancestry is overall beneficial, not deleterious, a similar pattern could also be produced from a number of different distributions of fitness effects for maize alleles, including for example if most maize alleles are deleterious but some have strong beneficial consequences. While maize ancestry in general is not predicted to provide adaptive benefits in teosinte, invasive *mexicana* in Europe shows selective sweeps for maize ancestry at multiple loci that have contributed to its establishment as a noxious weed [50] and we speculate that maize could be a source of alleles adapted to human-modified landscapes.

We repeated these analyses using local ancestry calls as our introgression estimates and found a non-significant Spearman's rank correlation between *mexicana* ancestry and recombination rates for 1 cM windows in sympatric maize ($\rho = 0.011,\ \mathrm{CI}_{95}[-0.039,\ 0.062]$) and a negative rank correlation between *mexicana* ancestry and recombination rate in sympatric *mexicana* ($\rho = -0.473,\ \mathrm{CI}_{95}[-0.512,\ -0.432]$). Contrasting results between global and local ancestry methods could be a reflection of true evolutionary differences across different time periods; local ancestry methods capture patterns from more recent gene flow that comes in longer ancestry blocks while STRUCTURE-like algorithms (NGSAdmix) and f_4 statistics are based on allele frequencies that collapse information across ancestry blocks of any size, capturing a longer evolutionary time scale. This interpretation would suggest that *mexicana* has experienced stronger selection against more recent maize gene flow than historical gene flow. However, we caution that local ancestry methods may

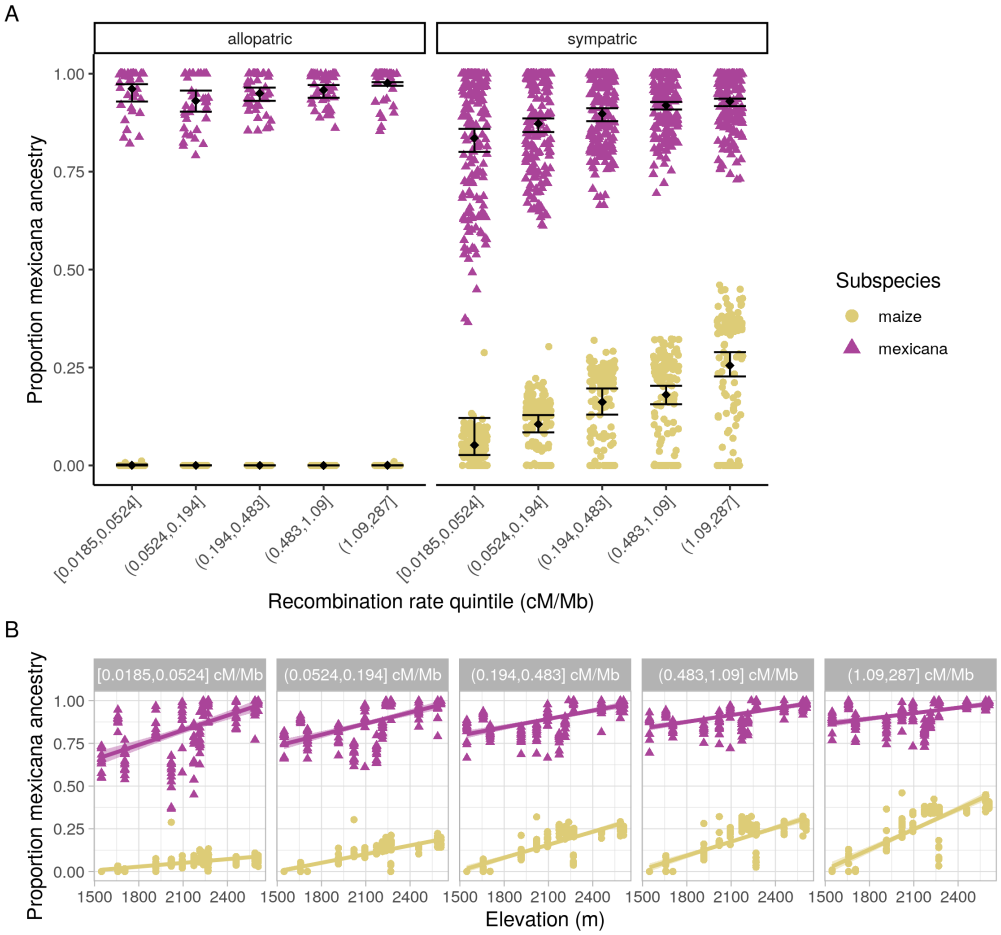

Fig 4. (A) *Mexicana* ancestry by recombination rate. Inferred *mexicana* ancestry in allopatric reference populations (left) and sympatric maize and *mexicana* populations (right) using NGSAdmix (K=2) by recombination rate quintiles. Group mean and 95% confidence interval based on bootstrap percentiles (n = 100) are depicted in black. Ancestry estimates for each individual are shown as points, colored by subspecies, and points are jittered for better visualization. (B) Slopes of *mexicana* ancestry across elevation for each recombination rate quintile, based on NGSAdmix estimates. Each point is a sympatric maize or *mexicana* individual and lines show the best-fit linear model for ancestry by elevation (with shaded 95% confidence interval) estimated separately for each quintile and subspecies.

also have subtle biases in power that are sensitive to local recombination rates and make them less reliable for comparing ancestry patterns across recombination rate quintiles.

Overall, we find support for widespread selection against introgression into maize and mixed results from similar tests of this hypothesis in *mexicana*.

High introgression peaks shared across populations

To assess adaptive introgression in our sympatric populations, we identified introgression 'peaks' where minor ancestry exceeds the genomewide mean by more than 2 standard deviations. We find no strong reduction in average diversity (π) for *mexicana* ancestry at high introgression peaks (Fig S16). This maintenance of diversity

implies that selection at most peaks has favored multiple *mexicana* haplotypes, and hard sweeps for a recent beneficial mutations on a specific haplotype are rare.

We observe that many high *mexicana* ancestry peaks are shared across subsets of our 14 maize landrace populations (see e.g. chr4, Fig 5). While most outlier peaks are unique to a single population, many peaks are shared across 7 or more of the populations (Fig S13). To a lesser extent, we also observe sharing of high-introgression peaks for maize ancestry in sympatric *mexicana* populations (Fig S14).

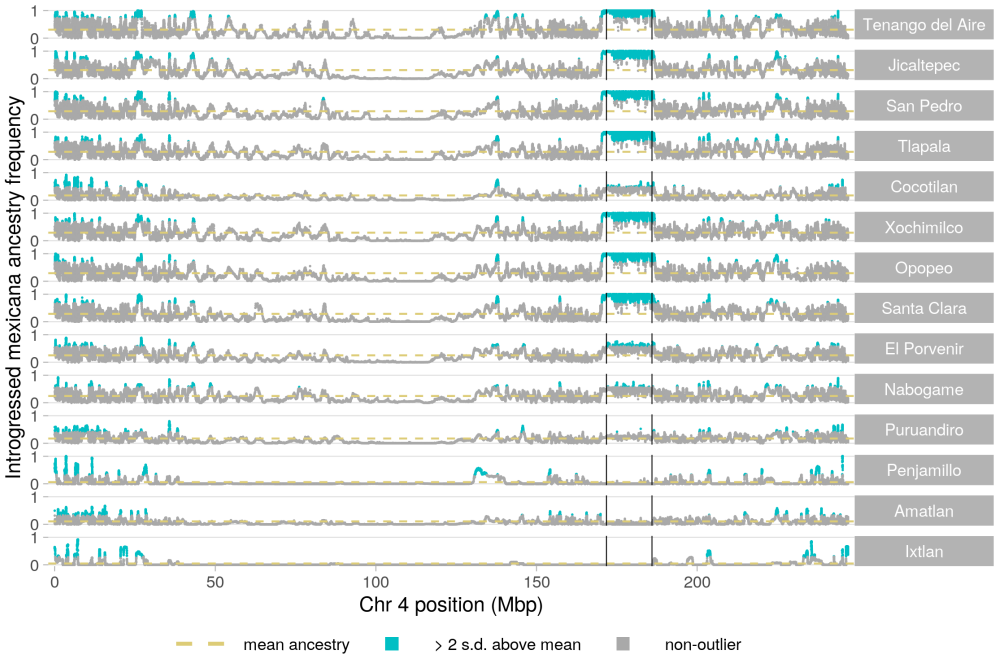

Fig 5. Introgression in maize landrace populations across chromosome 4. Local introgressed ancestry frequency for each maize landrace population compared to their genomewide mean. Populations are ordered from high to low elevation (top to bottom). High introgression peaks with more than 2 standard deviations above the population mean introgressed *mexicana* ancestry are highlighted in blue. Vertical black lines show the previously identified endpoints for a large inversion (*Inv4m* coordinates from Fig 3 of [79]). For local ancestry on other chromosomes, see Figs S20-S38
.

High introgression peaks in many independent populations would be very unexpected by chance. However, our sampled populations do not provide independent evidence for adaptive introgression, due to shared gene flow and drift post-admixture (e.g. long-distance human-assisted dispersal of maize seed). To estimate the rate of peak sharing we should expect from demographic processes alone, we simulated 100,000 unlinked loci under a multivariate normal distribution parameterized with the empirical ancestry variance-covariance matrix K (see methods). These simulations preserve the ancestry variance across loci within populations and non-independence in ancestry between populations.

For both sympatric maize and *mexicana*, every population shares an excess of high introgression peaks with all other populations compared to expectations set by our MVN null model. However, peak sharing is most elevated among high elevation maize populations (with the exception of Cocotilan, see Fig 6). To investigate the origins of population-specific peaks of introgression, we calculated F_{ST} between homozygous *mexicana* ancestry in local maize and in each *mexicana* population for these genomic

regions. Patterns of F_{ST} between local sympatric pairs at local introgression peaks differed little from background F_{ST} (Fig S17), offering little support for the idea that population-specific peaks arose from recent, locally sourced, adaptive introgression. Instead, patterns in maize are consistent with introgressed *mexicana* ancestry tracts from old shared admixture being favored by natural selection, and thus rising to high frequency, in a subset of landraces.

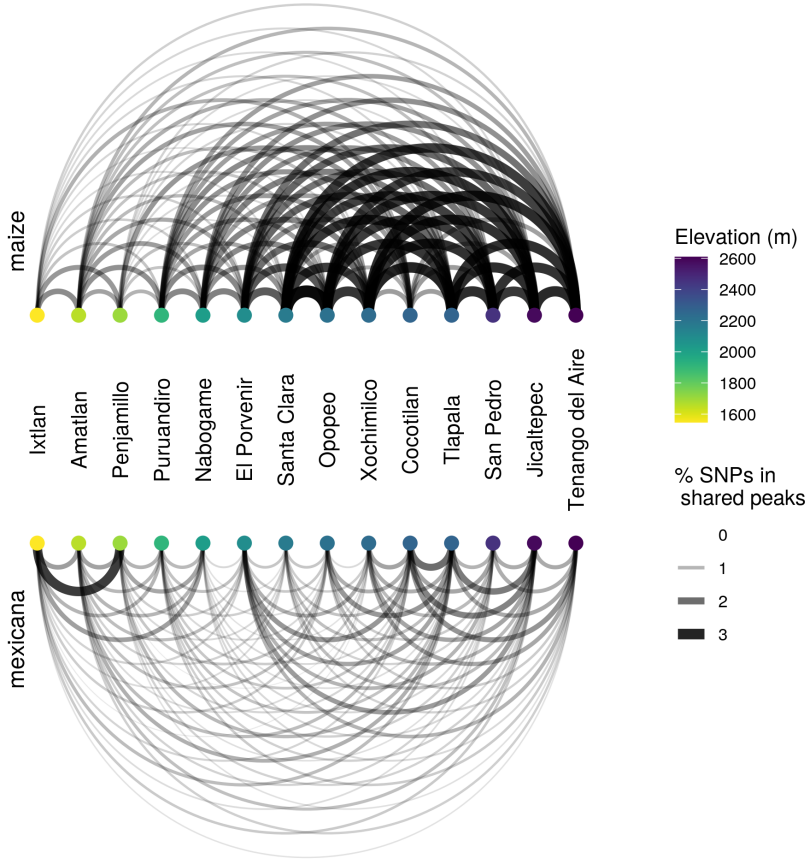

Fig 6. Introgression peaks shared across populations Networks for sympatric maize (top) and *mexicana* (bottom), where each node is a sampled population labelled by locationa and ordered by elevation. Edges connecting a pair of populations represent the percent of SNPs within shared ancestry peaks (introgressed ancestry > 2 s.d. above each population's mean ancestry). Sharing between all pairs of populations exceeds expectations based on multivariate-normal simulations that model genomewide covariance in ancestry.

This lack of local adaptive introgression is perhaps surprising given the genetic structure in *mexicana* associated with different ecotypes [29] and evidence for local adaptation within teosinte across elevation [31, 67, 79]. However, *mexicana* also has substantial standing variation and we find little evidence for hard sweeps, so one possibility is that local maize and local *mexicana* are are adapting to the same environment by different available genetic paths, or even the same causal SNP on a different set of haplotype backgrounds. Older introgressed tracts may also offer more accessible paths for maize adaptation, having already purged some of their linked deleterious variation. Additionally, local exitinction and re-colonization by *mexicana* is

common [109] and may contribute to a lack of local sourcing of adaptive haplotypes from contemporary *mexicana* populations.

Genomewide scan for selection on introgressed ancestry

We scanned the genome for two types of widespread selection on introgressed ancestry: consistent selection across populations creating an overall excess or deficit of introgression and fitness trade-offs creating steep clines in *mexicana* ancestry across elevation. We used our MVN simulated ancestry frequencies to set false-discovery-rates for excess and deficits of *mexicana* ancestry as well as steeper than expected slopes between *mexicana* ancestry and elevation.

We find several regions with high introgression in both directions that are unlikely to be explained by shared demographic history alone (Fig 7). These regions of adaptive introgression ($< 5\%$ FDR) are spread across the genome and cover a small fraction ($<0.5\%$) of the genome in both subspecies. We do not have power to determine if individual genes or regions are barriers to introgression because zero introgressed ancestry is not unusual under our simulated neutral model, given both low genomewide introgression and positive ancestry covariance between admixed populations (Fig 7).

Additionally, we identify outlier loci across the genome where *mexicana* ancestry forms steep clines across elevation (Fig 7). Our top candidate for strong associations between introgression and elevation in maize is *Inv4m*, a large 14 Mb inversion on chromosome 4 previously identified to have introgressed into high elevation maize landraces [15, 34, 42, 105]. This inversion maintains steep elevational clines within teosintes [79], overlaps QTLs for leaf pigmentation [49] and macrohairs [49], and is associated with increased yield in maize at high elevations and decreased yield at low elevations [15], but has thus far eluded functional characterization of genes within the inversion [15].

Our second strongest association co-localizes with *macrohairless1* (*mhl1*), a locus on chromosome 9 that controls macrohair initiation on the leaf blade [63] and is associated with a major QTL explaining 52% of macrohair variation between high and low elevation teosinte mapping parents [49]. Within teosintes, populations of the lowland ancestor of maize, *parviglumis*, show convergent soft sweeps at the *mhl1* locus not shared by *mexicana* [30]. Macrohairs are characteristic highland phenotypes in teosinte and maize and are thought to confer adaptive benefits through insect defence and/or heat maintenance [49, 64]. We identified a 3 Mb outlier region within the larger *mhl1* QTL which we analyzed further using PCA. We found three genetic clusters along the first principal component, evidence that an inversion polymorphism (hereafter *Inv9f*) maintains differentiation between maize/*parviglumis* and *mexicana* haplotypes across this region (Figs S18, S19). Additionally, principal component two at this inversion separates haplotypes genotyped in maize vs. those genotyped in *mexicana*, consistent with the *mexicana* allele at this inversion introgressing into maize long enough ago to accumulate maize-specific variation, and subsequently sorting in frequency across contemporary maize populations.

The clinal patterns of admixture that we observe at inversions *Inv4m* and *Inv9f* suggest they contribute to elevation-based adaptation in maize, with variation in their fitness impacts even within the historic elevational range of *mexicana*. While our highest peaks localize with regions previously associated with characteristic highland phenotypes, many additional outlier regions with steep increases in *mexicana* ancestry across elevation have undiscovered associations with local adaptation to elevation. Additionally, outliers for steep ancestry slopes across elevation in sympatric *mexicana* suggest that introgression from maize into *mexicana* may facilitate adaptation in *mexicana* to the lower end of its elevational range.

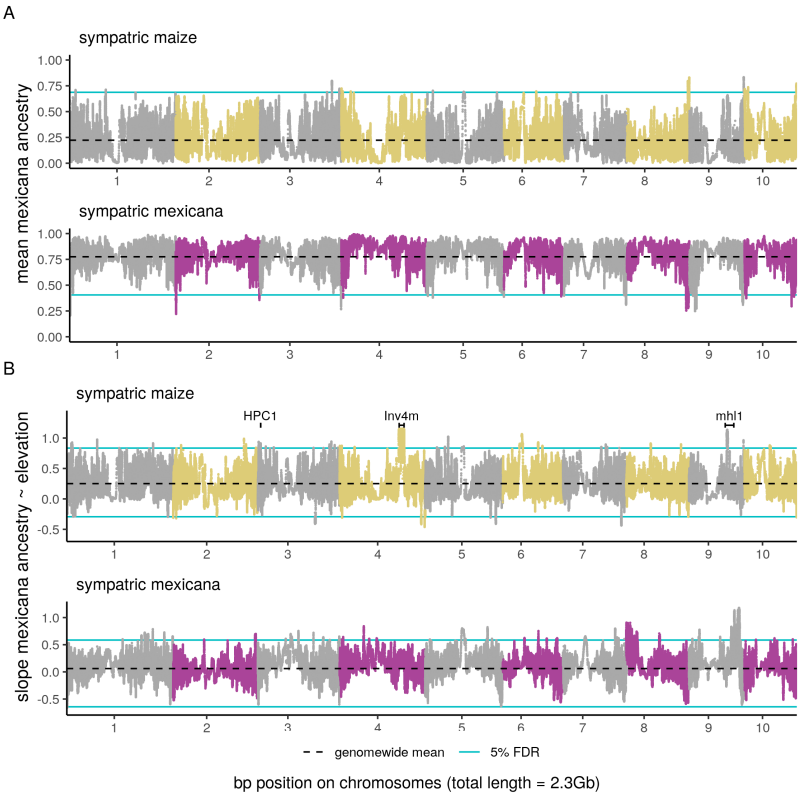

Fig 7. Genomewide scan for selection on *mexicana* ancestry. (A) Mean *mexicana* ancestry in sympatric maize and *mexicana* populations. (B) Slope of *mexicana* ancestry proportion over a 1 km elevation gain in sympatric maize and *mexicana* populations. In both (A) and (B) the blue lines shows the 5% false discovery rates, set using multi-variate normal simulations. Observing *mexicana* ancestry of 0% in sympatric maize or 100% in sympatric *mexicana* was not unexpected based on simulations. Positions for *Inv4m* [79] and the *mhl1* locus [63] were converted to v4 coordinates using Assembly Converter (ensembl.gramene.org). Chromosome numbers are placed at the centromere midpoint (approximate centromere positions are from [44]).

Selection at candidate domestication genes

We hypothesized that domestication genes will be barriers to introgression bilaterally between maize and *mexicana* [42]. While we do not have power to identify individual outlier genes that have low introgression, we can test for enriched overlap between 'introgression deserts' and a set of putative domestication genes spread across the genome.

We examined introgression for a sample of 15 well-characterized domestication genes from the literature (see S8 Table), and compared them to the regions of the genome with the lowest 5% introgression genomewide across all sympatric maize or *mexicana* populations ('introgression deserts'). A small but enriched subset of these domestication genes overlap with introgression deserts in sympatric maize ($N = 7$, $P < 0.001$) and likewise in sympatric *mexicana* ($N = 7$, $P < 0.001$). Among these candidates, we find that *teosinte branched1* (*tb1*), a key transciption factor that regulates branching vs. apical dominance [20, 21], overlaps introgression deserts in both maize and *mexicana*, consistent with *tb1*'s role at the top of the domestication regulatory hierarchy [23]. We also find evidence for reduced introgression into both maize and *mexicana* at *teosinte*

glume architecture1 (*tga1*) [24, 104] and *brittle endosperm2* (*bt2*) [107], which are associated with 'naked' edible grains and starch biosynthesis, respectively. Another eight domestication genes [55, 90, 93, 103, 107, 111, 112] have low introgression in one direction only. Among these, *sugary1* (*su1*) in the starch pathway has low maize ancestry in *mexicana* but shows a steep increase in introgressed *mexicana* ancestry with elevation in maize (< 5% FDR), which suggests this gene has pleiotropic effects on non-domestication traits in maize, with fitness trade-offs across elevation. *Sugary1* mutations affect seed germination and emergence at cold temperatures, candidate pleiotropic effects that could be more deleterious at higher elevations [99]. The remaining four domestication genes do not overlap introgression deserts in either subspecies despite evidence for their role in domestication: *zfl2* (cob rank) [9, 18, 19], *ba1* (plant architecture) [32], *ZmSh1-5.1+ZmSh1-5.2* (seed shattering) [55] and *pbf1* (storage protein synthesis) [106]. Despite evidence of introgression at many domestication loci, maize landraces retain all of the classic domestication traits, and *mexicana* populations maintain 'wild' forms. Epistasis for domestication traits [96] could help explain this discrepancy if compensatory effects from other loci contribute to maintaining domestication traits in admixed highland maize, or key domestication alleles segregate at moderate frequencies within *mexicana* (but do not have the same phenotypic effects in a teosinte background).

Selection within the flowering time pathway

Flowering earlier is adaptive in high-elevation environments where days are cooler and there are fewer total growing degree days in a season. We therefore expect an excess of introgressed *mexicana* ancestry at flowering time genes that may contribute to adaptive early flowering in highland maize. The *mexicana* allele at *High PhosphatidylCholine 1* (*HPC1*) has been shown to reduce days to flowering, confers a fitness benefit in maize at higher elevations, and is introgressed in modern flint maize cultivars from Northern Europe and America [84]. Here we show that *HPC1* also overlaps an outlier region with one of the steepest increases in *mexicana* ancestry with elevation across sympatric maize populations (+0.91 *mexicana* ancestry proportion/km, FDR < 5%), consistent with a role in adaptive earlier flowering at higher elevations in Mexican landraces. *HPC1* has low introgression from maize into sympatric *mexicana* across the elevational range of this study, suggesting either that the high-elevation *mexicana* allele does not confer fitness tradeoffs in the teosinte background or alternative segregating *mexicana* alleles maintain fitness at lower elevations.

We also tested for selection within the flowering time pathway more broadly using a set of candidate flowering time genes [22, 54] (N = 849). Only 1/43 genes from the core flowering time pathway (*ZMM5*) [22] and 15/806 other candidate flowering time genes [54] (+/- 20kb) overlap outlier regions with steep increases in *mexicana* introgression with increasing elevation (< 5% FDR) in sympatric maize, which matches expected overlap by chance (∼2%, P = 0.76). Thus the steep clinal introgression pattern at *HPC1* in sympatric maize, indicative of strong fitness trade-offs across elevation, is the exception, not the rule, for flowering-time related genes. While for *HPCI* the effect of the *mexicana* allele on reducing flowering time has been confirmed by CRISPR [84], a limitation for other less-characterized genes is that we simply assume the *mexicana* allele reduces flowering time. In addition, flowering time is a highly polygenic trait [11], which could reduce the strength of selection at individual genes with smaller effect sizes than *HPC1* to below what we can detect using steep ancestry clines. While it is alternatively possible that *mexicana* alleles would show adaptive benefits across the entire range sampled (moderate to high elevation), we find that only 2 candidate flowering time genes overlap high mean *mexicana* introgression outliers at a 5% FDR (P = 0.23).

Conclusion

We conclude that the majority of *mexicana* ancestry introgressed into maize over 1000 generations ago and has subsequently been sorted across an elevational gradient, and by selection within individual populations. Differentiation of *mexicana* haplotypes within maize genomewide (F_{ST}) and at individual introgressed outlier loci (e.g. *Inv9f* at the *mhl1* locus (PCA)) corroborate this timeline. Despite contemporary observations of hybrids in the field [109], there is little evidence of significant recent gene flow in either direction between sympatric maize-*mexicana* population pairs. Consistent with domestication loci acting as barriers to introgression, in both maize and *mexicana* an enriched subset of candidate domestication genes overlap 'introgression deserts.' More generally, we find introgressed *mexicana* alleles are on average deleterious in maize, but less evidence for a genomewide effect of selection against introgression into *mexicana*, possibly because epistasis masks the impact of maize alleles in a *mexicana* background [96]. Some loci show exceptional ancestry patterns consistent with selection favoring introgression in multiple populations, especially for *mexicana* ancestry in the highest elevation maize. While these shared signatures of adaptive introgression are the most striking, the majority of ancestry peaks are exclusive to a single population. Despite this signature of geographically-restricted local adaptation from *mexicana* ancestry, there is no evidence of local population sources for locally adapted haplotypes at these peaks. Thus both broad and local adaptation of maize throughout the highlands appears to have been driven primarily by the sorting of old introgression.

Materials and Methods

Population sampling

We used maize and *mexicana* seed accessions sampled from locations across Mexico in 2008 [42] and currently stored at UC Davis. We included 14 maize and 14 *mexicana* accessions that are paired populations sampled in sympatry from the same locations: Ixtlan*, Amatlan, Penjamillo, Puruandiro*, Nabogame*, El Porvenir*, Santa Clara*, Opopeo*, Xochimilco*, Cocotilan, Tlapala, San Pedro*, Jicaltepec and Tenango del Aire* (see S1 Table). A previous study of crop-wild admixture genotyped different maize and *mexicana* individuals from 9 of these locations (marked with *), using the Illumina MaizeSNP50 Genotyping BeadChip [42]. In addition, we chose three *mexicana* population accessions with no record of contemporary maize agriculture in the vicinity for an allopatric *mexicana* reference panel: Puerta Encantada, Malinalco, and Amecameca.

At each sampling location, multiple ears from maternal plants were collected for seed. Population accessions varied in the number of maternal plants with viable seeds. When available, we planted multiple seeds within each ear but only randomly selected one individual for sequencing from the plants that successfully germinated in the greenhouse.

DNA extraction and sequencing

We extracted DNA from leaf tissue and then prepared sequencing DNA libraries using a recently published high-throughput protocol ("Nextera Low Input, Transposase Enabled protocol" [86]) with four main steps: (1) DNA shearing and tagmentation by the Nextera TD enzyme, (2) PCR amplification (Kapa2G Robust PCR kit) and individual sample barcoding (custom 9bp P7 indexing primers) (3) library normalization and pooling, and (4) bead-based clean-up and size-selection of pooled libraries. We sequenced the resulting pooled libraries using multiple lanes on Illumina HiSeq 4000 and Novaseq 6000 machines (paired-end 150 bp reads).

To address low sequencing output from some libraries, we re-sequenced 26 libraries (and merged output) and replaced 53 lower-coverage libraries with a higher-coverage library prepared from another seed grown from the same half-sibling family. We excluded 7 samples from analysis because their final libraries did not yield sufficient sequencing output (<0.05x coverage after filtering reads for mapping quality). We additionally removed one lane of sequencing (58 samples) from the study after determining a labelling error had occurred for that plate. In total, we obtained whole genome sequences for 348 individuals (1.0x average coverage, range: 0.1-2.4x). Of these samples, 43 are *mexicana* from three allopatric populations, with a total of 34.1x combined coverage. The remaining samples are maize and *mexicana* from sympatric populations, 262 of which have sufficient coverage for local ancestry inference ($\geq$ 0.5x, 6-12 per sympatric population, see Fig S1). Raw sequencing reads for these low-coverage maize and *mexicana* genomes are available at NCBI (PRJNA657016).

Reference genome and recombination map

We used version 4 of the B73 maize reference genome [44] (Zea_mays.B73_RefGen_v4.dna.toplevel.fa.gz, downloaded 12.18.2018 from Gramene).

To find local recombination rates, we converted marker coordinates from a published 0.2 cM genetic map [66] to the v4 maize genome using Assembly Converter (ensembl.gramene.org). We removed any markers that mapped to a different chromosome or out of order on the new assembly, and extended the recombination rate estimates for the most distal mapped windows to the ends of each chromosome. From this map, we used *approx()* in R (v3.6.2 [81]) to estimate the cM position for any bp position, based on linear interpolation.

Read mapping and filtering

First, we checked read quality using fastQ Screen (v0.14.0 [113]) and trimmed out adapter content from raw sequencing reads using the trimmomatic wrapper for snakemake (0.59.1/bio/trimmomatric/pe) [8]. We mapped trimmed reads to the maize reference genome using bwa mem (v0.7.17 [52]). We then sorted reads using SAMtools (v1.9 [53]), removed duplicates using picardtools (v2.7.1) MarkDuplicates and merged libraries of the same individual sequenced on multiple lanes using SAMtools merge. In all subsequent analyses in the methods below we filtered out reads with low mapping scores (< 30) and bases with low base quality scores (< 20).

High-coverage *Tripsacum* genome sequencing

In addition to low-coverage genomes for maize and *mexicana*, we selected a *Tripsacum dactyloides* individual as an outgroup and sequenced it to high coverage. This individual is an outbred 'Pete' cultivar (rootstock acquired from the Tallgrass Prairie Center, Iowa, USA). We extracted genomic DNA from leaf tissue using the E.Z.N.A.® Plant DNA Kit (Omega Biotek), following manufacturer's instructions, and then quantified DNA using Qubit (Life Technologies). We prepared a PCR-free Truseq DNA library and sequenced it with an Illumina HiSeq2500 rapid run (paired-end 250 bp reads). We generated a total of 136.53 Gb of sequencing for this individual, available at NCBI (SRR7758238). For the following analyses that use *Tripsacum* as an outgroup, we randomly subsampled 50% of reads using seqtk, for approximately 30x coverage. We mapped reads to the maize reference using the pipeline described above, and additionally capped base quality scores with the 'extended BAQ' model in SAMtools [51], which reduces the influence of bases with lower alignment quality.

Additional genomes from published sources

For an allopatric maize reference population, we used 55 previously published high-coverage maize genomes from Palmar Chico (NCBI: PRJNA616247 [13]). This landrace population grows below the elevational range for *mexicana* (1,008 m).

For a parviglumis reference population, we used 50 previously published high-coverage lowland individuals sampled from the 'Mound' population at 1,008 m by Palmar Chico [13, 101, 114] (NCBI: PRJNA616247, see S2 Table). We mapped and filtered reads for these individuals using the pipeline described above and capped base quality scores using BAQ.

SNP calling

We called SNPs using a combined panel of the 348 low-coverage genomes sequenced in this study for sympatric maize, sympatric *mexicana*, and allopatric *mexicana*, and the 55 high-coverage allopatric maize reference genomes described above. We used ANGSD (v0.932 [47]) to identify variant sites with minor allele frequencies $\geq 5\%$ in the total sample based on read counts ('angsd -doMajorMinor 2 -minMaf 0.05 -doCounts 1 -doMaf 8'). In addition to mapping and base quality filters ('-minMapQ 30 -minQ 20'), we capped base qualities using the extended per-Base Alignment Quality algorithm ('-baq 2' [51]) and removed sites that did not have at least 150 individuals with data or had sequencing depth exceeding 2.5x the total sample mean depth. To apply this total depth filter, we estimated mean depth ('angsd -doCounts 1 -doDepth 1 -maxDepth 10000') for 1000 regions of length 100bp randomly sampled using bedtools (v2.29.0 [80]). In total, we identified 52,118,357 SNPs on the assembled chromosomes. In conjunction with SNP calling, we produced beagle-format genotype likelihoods for each individual at these variant sites using the SAMtools GL method [53] implemented in ANGSD ('-GL 1 -doGlf 2').

Global ancestry inference

To estimate genetic relationships between populations and genomewide ancestry proportions, we used methods specific to low-coverage data that rely on genotype likelihoods, rather than called genotypes. Because these methods are sensitive to SNPs in high linkage disequilibrium (LD), we thinned genotype likelihoods to every 100th SNP ($\sim$4kb spacing) [98]. To confirm that maize and *mexicana* subspecies form a major axis of genetic variation in our sample, we estimated the genetic covariance matrix between individuals using PCAngsd (v0.98.2 [60]) and visualized principal components computed using *eigen()* in R. We then estimated global ancestry proportions using the same thinned genotype likelihood files as input to NGSAdmix [91], using K = 2 clusters. Clusters clearly mapped onto the allopatric reference groups, which we used to label the two ancestry components as 'maize' and '*mexicana*'.

Local ancestry and timing of admixture

We inferred local ancestry across the genome using a hidden Markov model that is appropriate for low-coverage data because it models genotype uncertainty down to the level of read counts for all admixed individuals (ancestry_hmm [14]). This method relies on allele counts from separate allopatric populations to estimate reference population allele frequencies for each ancestry. Because some of our allopatric reference individuals have too low of coverage to accurately call genotypes, we randomly sampled one read per individual to get unbiased frequency estimates for major and minor alleles at each site ('angsd -doCounts 1 -dumpCounts 3'). To maximize ancestry-informativeness of

sites in this analysis, we identified SNPs with allele frequency differences of at least 0.3 between reference populations ('angsd -doMajorMinor 3 -GL 1') based on at least 44 allopatric maize and 12 allopatric *mexicana* individuals having coverage at a site. We then calculated genetic distances between SNPs and filtered our enriched variants down to 1 SNP per 0.01 cM using the maize recombination map.

Running ancestry_hmm jointly infers local ancestry for each individual and the time since admixture. Each sympatric maize or *mexicana* population was analysed separately, using the population's mean NGSAdmix global ancestry estimate as a prior for mixing proportions, 100 generations as a prior for admixture time (range: 0-10000), an approximate effective population size (Ne) of 10,000 individuals and an estimated sequencing base error rate of 3×10^{-3}. We ran ancestry_hmm with an optional setting to bootstrap 100 random samples of 1,000-SNP genomic blocks to estimate uncertainty around the estimated generations since admixture. To test the sensitivity of the HMM to our choice of Ne, we re-ran ancestry_hmm with two other Ne's that differ by an order of magnitude (Ne = 1k, 100k), but did not analyze these results further after finding high correspondence for both local ancestry and timing estimates.

Local ancestry output is in the form of a posterior probability for homozygous maize, homozygous *mexicana* or heterozygous ancestry for each individual at every site. To get a single point estimate for local ancestry at a site for an individual, we summed *mexicana* ancestry, weighted by these posterior probabilities: *mexicana* ancestry proportion = P(homozygous *mexicana*) + $^1/_2$P(heterozygous). In addition, for analyses that require ancestry tract positions, we assumed that the estimated ancestry at a focal site extends halfway to the next site with a local ancestry estimate.

Diversity within ancestry

Using *mexicana* ancestry estimates from the HMM, we identified high-confidence homozygous ancestry tracts for both maize and *mexicana* ancestry (posterior > 0.8). We filtered individual bams for reads that overlap these tracks and used the resulting filtered bams to calculate diversity within both maize and *mexicana* ancestry, separately. We estimate diversity using the ANGSD/realSFS framework which is appropriate for low-coverage sequence data it takes into account uncertainty in both genotypes and variant sites. We created a concensus fasta sequence for *Tripsacum* ('angsd -doFasta 2') to use as the ancestral state for these analyses.

For each population and ancestry, we estimated the site allele frequencies ('angsd -doSaf 1 -GL 1') and subsequently estimated the genomewide site frequency spectrum (SFS). We then used this SFS as a prior to estimate within-ancestry pairwise diversity (π) genomewide from the site allele frequencies ('realSFS saf2theta').

For each pair of populations and ancestry, we additionally used realSFS to estimate the two dimensional SFS from the individual population site allele frequencies genomewide. We then used this 2D SFS as a prior to estimate genomewide within-ancestry F_{ST} between the two populations ('realSFS fst index -whichFst 1'). This call uses Hudson's F_{ST} estimator [40] as parameterized in [7].

Effect of local recombination rate on introgressed ancestry

To estimate the effects of linked selection and recombination rate on genomewide introgression patterns, we compared *mexicana* ancestry estimates across genomic quintiles. Based on a 0.2 cM-resolution recombination map [66] for maize, we merged adjacent recombination windows into larger 1 cM non-overlapping windows and calculated each window's mean recombination rate and overlap with coding base pairs (bedr 'coverage') [37]. We retrieved coding regions ('CDS') using gene annotations from

Ensembl (ftp://ftp.ensemblgenomes.org/pub/plants/release-41/gff3/zea_mays/Zea_mays.B73_RefGen_v4.41.chr.gff3.gz, dowloaded 11.6.2018). We sorted windows into quintiles for either recombination rate or coding density (bp/cM). Each quintile covers approximately 1/5 of the genome based on physical bp.

i. NGSAdmix estimates

To estimate ancestry proportions for each recombination rate quintile, we first reduced LD by thinning to 1% of SNPs (every 100th) and ran NGSAdmix 5 times separately (once per quintile) with K=2 clusters. We assigned 'maize' and '*mexicana*' labels to the ancestry clusters based on majority assignment to the respective allopatric reference panels. To bootstrap for uncertainty, we re-sampled 1 cM windows with replacement from each quintile 100 times, and re-ran NGSAdmix on the resulting bootstrap SNP sets. Using these results, we calculated 95% percentile bootstrap confidence intervals for the estimated admixture proportions, and the Spearman's rank correlation between the recombination rate (or coding bp per cM) and admixture proportion ranks for each quintile. We also tested for a difference in ancestry slopes with elevation across different recombination rate quintiles by fitting a linear model with an elevation by recombination quintile interaction term: *mexicana* ancestry $\sim$ elevation + r + elevation*r. Using lm() in R, we fit this model for sympatric maize and sympatric *mexicana* separately, treating quintiles as a numeric scale 0-4.

f_4 estimates

In a complementary analysis, we used a ratio of f_4 statistics as an alternative method to estimate ancestry proportions by quintile. Assuming the basic phylogenetic tree (((*parviglumis*, allopatric maize), allopatric *mexicana*), *Tripsacum*) in Fig S5, we can estimate α, the proportion of ancestry inherited from *mexicana* in an admixed population, as follows [36, 70]:

$$\alpha = \frac{f_4(\text{Tripsacum, } \textit{parviglumis}\text{; X, allopatric maize})}{f_4(\text{Tripsacum, } \textit{parviglumis}\text{; allopatric } \textit{mexicana}\text{, allopatric maize})}.$$

The denominator of this statistic estimates the branch length leading to *parviglumis* and allopatric maize that separates these sister subspecies from allopatric *mexicana*; the full ratio estimates the proportion of this branch that separates sympatric population X from *parviglumis* and allopatric maize, i.e. the *mexicana* ancestry in X. Because the f_4 statistic is sensitive to additional unmodeled admixture within the tree, we limited our allopatric *mexicana* group to individuals from just one of the three reference populations (Amecameca), which showed no evidence of admixture in our global ancestry analysis (see Fig 2).

For each 1 cM window across the genome, we used ANGSD to calculate ABBA-BABA statistics from observed read counts for the 4 populations in the numerator and denominator of the α estimator separately ('angsd -doabbababa2 1 -remove_bads 1 -minMapQ 30 -minQ 20 -doCounts 1 -doDepth 1 -maxDepth 10000 -useLast 1 -blockSize 5000000'). From the resulting output files, we summed the negative of the ABBA-BABA numerator ('Num') and divided by the total number of included sites ('nSites') across all 1 cM windows within a quintile to get the f_4 statistic [92].

We then calculated the Spearman's rank correlation between the recombination rate quintiles and admixture proportion ranks for these quintiles. We calculated simple bootstrap confidence intervals for our ancestry estimates and correlations by re-sampling 1 cM windows within quintiles with replacement 10,000 times and re-calculating the f_4 ratios and resulting rank correlation across quintiles to construct

95% percentile confidence intervals. We repeated this analysis using quintiles based on coding bp per cM in place of recombination rate (cM/Mbp).

ii. Local ancestry estimates

We also calculated the Spearman's rank correlation between local recombination rate (or coding bp per cM) and local ancestry proportion at the level of individual 1 cM windows. For each window, we averaged local ancestry estimates from the HMM across all individuals within sympatric maize, and separately, sympatric *mexicana*. We then calculated simple bootstrap confidence intervals for our local ancestry estimates and local recombination rate (or coding bp per cM) by re-sampling 1 cM windows across the genome with replacement 10,000 times and re-calculating the rank correlation across windows to construct 95% percentile confidence intervals.

Local ancestry simulations

We simulated *mexicana* ancestry population frequencies using a multivariate-normal null model:

$$\textit{mexicana} \text{ ancestry} \sim \text{MVN}(\vec{\alpha}, K)$$

where $\vec{\alpha}$ is the vector of mean *mexicana* ancestry frequencies genomewide for each sympatric population and K is the empirical ancestry variance-covariance matrix relating these 14 populations. The diagonal entries of the K matrix capture the expected variation in local ancestry across the genome within populations due to drift and random sampling. The off-diagonals capture ancestry covariances between populations created by shared gene flow and drift post-admixture: at loci where one population has an excess of introgression, other admixed populations with shared demographic history will also tend to have an excess of introgression.

To construct K, we calculated the covariance in ancestry between each pair of populations i and j using all L loci with local ancestry calls genomewide:

$$K[i,j] = \frac{1}{L}\sum_{l=1}^{L}(Anc_{i,l} - \alpha_i)(Anc_{j,l} - \alpha_j).$$

Above, $Anc_{i,l}$ and $Anc_{j,l}$ are local ancestry frequencies at a locus l while α_i and α_j are the mean local ancestry frequencies across the genome for populations i and j.

For sympatric maize and sympatric *mexicana* separately, we calculated the empirical K matrix between populations from all 14 sympatric locations, and then took 100,000 independent draws from their MVN distribution, thereby simulating *mexicana* ancestry for all populations at 100,000 unlinked loci. Because ancestry frequencies are bounded at [0,1] but normal distributions are not, we truncated any simulated values outside of this range.

Introgression peaks shared between populations

To characterize introgression peak sharing between individual populations, we defined 'ancestry peaks' as sites where a population has over 2 standard deviations more introgressed ancestry than the genomewide mean. We counted the number of peaks that are shared between all pairs and combinations of populations. To compare these results to our null model, we also counted the number of introgression peaks shared by populations in our simulated dataset, using the 2 s.d. cutoff set by the empirical data to define peaks.

Because *mexicana* ancestry shows significant diversity, we additionally characterized diversity for *mexicana* ancestry peaks introgressed into maize. For all introgressed ancestry outlier regions in a focal maize population, we used ANGSD to estimate pairwise diversity within the population (π) and differentiation (F_{ST}) between the focal sympatric maize populations and their local sympatric *mexicana* population. We focused on the *mexicana* ancestry within peaks by limiting our diversity estimates to only include high-confidence homozygous *mexicana* ancestry tracts (posterior > 0.8). For these analyses, we pooled information across outlier peaks, but distinguish between introgression peaks exclusive to the focal population and introgression peaks shared between the focal population and at least 3 other sympatric maize populations. We used global estimates of the SFS and 2D SFS as priors to estimate π and F_{ST} for the subsets of the genome within introgression peaks, and otherwise followed the same methods listed above in 'Diversity within ancestry'.

Genomewide scan for ancestry outliers

For sympatric maize and *mexicana* separately, we calculated the mean *mexicana* ancestry across all individuals at a locus, and fit a linear model using lm() in R to estimate the slope of *mexicana* ancestry frequencies for sympatric populations across elevation (km): *mexicana* ancestry $\sim$ elevation. We then repeated these summary statistics for every locus with an ancestry call in the empirical data and each simulated locus in the MVN simulated data.

We calculated 5% false-discovery-rate (FDR) cutoffs for high and low *mexicana* ancestry using the Benjamini-Hochberg method [6] and simulation results to estimate the expected frequency of false-positives under our null MVN model (one-tailed tests). We repeated this approach to identify outlier loci with steep positive (or negative) slopes for *mexicana* ancestry across elevation at a 5% FDR.

Test for reduced introgression at domestication genes

To test whether domestication genes are unusually resistant to introgression, we first defined 'introgression deserts' as regions with the lowest 5% of introgression genomewide across all sympatric maize (or, separately, sympatric *mexicana*) populations. We then looked up v4 coordinates on Ensembl.org for genes associated with maize domestication in the literature (S8 Table), and used bedtools 'intersect' to identify which of these genes ± 20 kb overlap introgression deserts. To test for significance, we randomly shuffled the gene positions across the genome (bedtools 'shuffle') 1000 times and re-calculated overlap with introgression deserts for each permuted data set.

Test for selection within the flowering time pathway

We identified a list of 48 core flowering time pathway genes from the literature [22], and a broader list of 905 flowering time candidate genes [22, 54]. From the combined set, we included 849 total genes (43 core pathway) which we were able to localize on assembled autosomes of the v4 reference genome using MaizeGDB gene cross-reference files [75]. We counted the number of genes ± 20 kb that intersected with outlier regions for steep increases in *mexicana* introgression with elevation (and, separately, high *mexicana* introgression) in sympatric maize populations ($< 5\%$ FDR) using bedtools 'intersect', then tested for significance by repeating this analysis with 1000 randomly shuffled gene positions.

Analysis pipeline and data visualization

We constructed and ran bioinformatics pipelines using snakemake (v.5.17.0 [48]) within a python conda environment (v3.6). We analyzed and visualized data in R (v3.6.2 [81]) using the following major packages: tidyverse (v1.3.0 [108]), viridis (v0.5.1 [33]), bedr (v1.0.7 [37]), boot (v.1.3_25 [12,16]), gridExtra (v2.3 [4]), ggupset (v0.3.0 [3]) and tidygraph (1.2.0 [69]). All scripts can be found on our gitHub repository, https://github.com/ecalfee/hilo, which also includes a full list of software and versions (see envs/environment.yaml).

Supporting Information

S1 Table. Population metadata.

S2 Table. Parviglumis SRA IDs.

S3 Table. Ancestry by elevation and recombination rate quintile. Best-fitting linear models for ancestry proportion predicted by an elevation by recombination rate interaction: *mexicana* ancestry $\sim$ elevation + r + elevation*r. Here, r is the recombination rate quintile, treated as numeric [0-4]. This model only uses ancestry estimates for sympatric individuals and is fit separately for maize and *mexicana* samples.

group	term	estimate	std.error	statistic	p.value
sympatric maize	intercept	-0.122	0.027	-4.512	7.67E-06
sympatric maize	elevation (km)	0.083	0.013	6.596	9.01E-11
sympatric maize	r quintile	-0.110	0.011	-9.896	1.50E-21
sympatric maize	elevation*r quintile	0.074	0.005	14.399	8.27E-41
sympatric mexicana	intercept	0.270	0.035	7.710	3.35E-14
sympatric mexicana	elevation (km)	0.270	0.016	16.590	4.52E-54
sympatric mexicana	r quintile	0.119	0.014	8.309	3.57E-16
sympatric mexicana	elevation*r quintile	-0.045	0.007	-6.735	2.94E-11

S4 Table. Spearman's rank correlation between *mexicana* ancestry (NGSAdmix) and recombination rate (or coding bp per cM) quintiles

group	feature	Spearman's ρ	2.5%	97.5%
allopatric maize	recombination rate (cM/Mb)	-0.36	-0.90	0.67
allopatric mexicana	recombination rate (cM/Mb)	0.40	0.10	1.00
sympatric maize	recombination rate (cM/Mb)	1.00	0.85	1.00
sympatric mexicana	recombination rate (cM/Mb)	1.00	0.90	1.00
allopatric maize	coding bp per cM	0.62	-0.30	1.00
allopatric mexicana	coding bp per cM	-0.40	-0.90	-0.20
sympatric maize	coding bp per cM	-1.00	-1.00	-0.85
sympatric mexicana	coding bp per cM	-0.90	-1.00	-0.80

S5 Table. Spearman's rank correlation between ancestry (f_4 ratio) and recombination rate (or coding bp per cM) quintiles in sympatric maize

group	feature	Spearman's ρ	2.5%	97.5%
sympatric maize	recombination rate (cM/Mb)	1.00	0.30	1.00
sympatric maize	coding bp per cM	-0.90	-1.00	0.10

S6 Table. Spearman's rank correlation between ancestry (f_4 ratio) and recombination rate (or coding bp per cM) quintiles in sympatric *mexicana*

group	feature	Spearman's ρ	2.5%	97.5%
sympatric mexicana	recombination rate (cM/Mb)	1.00	0.80	1.00
sympatric mexicana	coding bp per cM	-1.00	-1.00	-0.90

S7 Table. Spearman's rank correlation between mean *mexicana* local ancestry and recombination rate (or coding bp per cM) at 1 cM genomic window resolution. Confidence intervals are constructed using the percentile method and 10,000 bootstrap replicates created by randomly re-sampling 1 cM windows within quintiles.

group	feature	Spearman's ρ	2.5%	97.5%
sympatric_maize	recombination rate (cM/Mb)	0.011	-0.039	0.062
sympatric_maize	coding bp per cM	0.014	-0.038	0.066
sympatric_mexicana	recombination rate (cM/Mb)	-0.473	-0.512	-0.432
sympatric_mexicana	coding bp per cM	0.339	0.292	0.384

S8 Table. Domestication genes and overlap with low introgression regions.

gene	phenotype	refs	v4 coordinates	min introgression in maize	min introgression in mexicana
zagl1	ear size	[111]	1:4959131-5014850	0.012*	0.264
gt1	prolificacy	[112]	1:23605801-23647370	0.004*	0.106
ZmSh1-1	seed shattering	[55]	1:228660490-228705551	0.356	0.038*
tb1	branching	[20, 21, 23]	1:270533676-270574776	0.001*	0.06*
zfl2	cob rank	[9, 18, 19]	2:12894091-12937068	0.113	0.271
pbf1	storage protein synthesis	[106]	2:158122366-158176919	0.092	0.084
ra2	inflorescence architecture	[103]	3:12138280-12179065	0.144	0.046*
ba1	plant architecture	[32]	3:185994629-186035264	0.33	0.154
su1	starch biosynthesis	[107]	4:43090569-43139167	0.445	0.017*
tga1	'naked' grains	[24, 104]	4:46330597-46375118	0.036*	0.007*
bt2	starch biosynthesis	[107]	4:61295575-61341350	0.027*	0.016*
ZmSh1-5.1+ZmSh1-5.2	seed shattering	[55]	5:16630307-16676707	0.105	0.175
sweet4c	sugar transport and seed size	[93]	5:130767030-130809864	0.014*	0.163
ae1	starch biosynthesis	[107]	5:172392995-172450415	0.436	0.059*
ra1	inflorescence architecture	[90, 103]	7:113552410-113592937	0.006*	0.176

* lowest 5% introgression genomewide ('introgression desert')

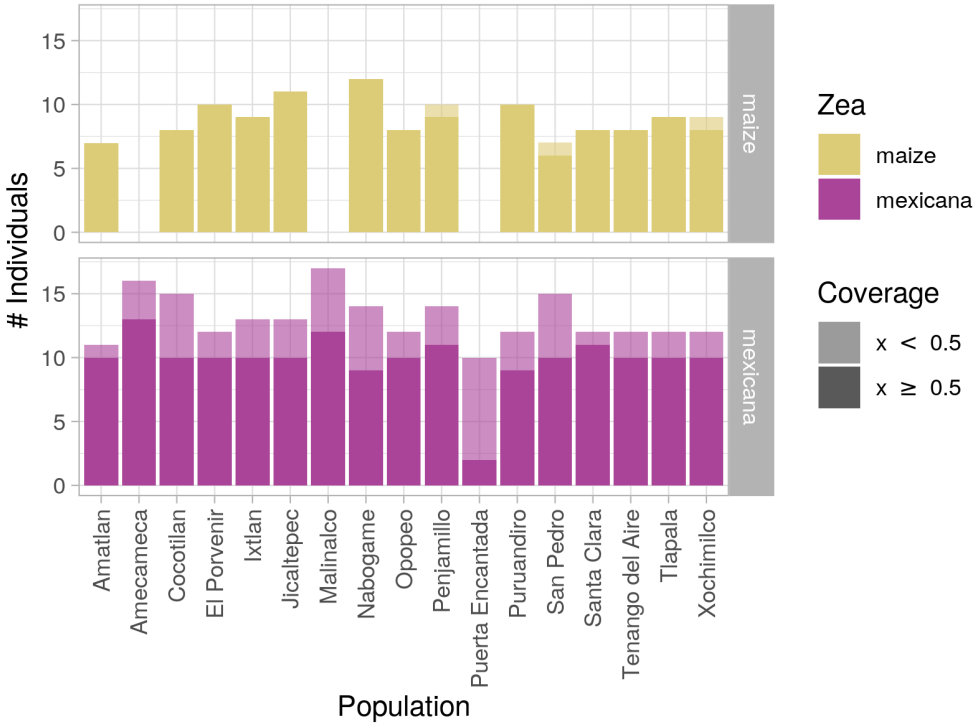

Fig S1. Number of individuals sequenced per location. Number of maize (top) and *mexicana* (bottom) individuals sequenced by this study with minimum 0.05x WGS coverage. Amecameca, Malinalco and Puerta Encantada have no paired maize samples. These *mexicana* populations were sampled from locations with no recent history of maize agriculture and are used as a reference panel for allopatric *mexicana* ancestry. For sympatric maize and *mexicana*, only individuals meeting a more stringent 0.5x coverage threshold (shown in darker shading) are included in analyses based on local ancestry inference.

Acknowledgments

The authors would like to acknowledge funding from NSF award number 1546719 to JRI and GC. This work was also supported by the National Institute of General Medical Sciences of the National Institutes of Health (NIH R01 GM108779 and R35 GM136290, awarded to GC). We thank Pesach Lubinsky for collecting the seeds sequenced in this study. We also want to thank the Coop and Ross-Ibarra labs, and the HILO and Zeavolution working groups for helpful feedback on this work.

References

1. S. Aeschbacher, J. P. Selby, J. H. Willis, and G. Coop. Population-genomic inference of the strength and timing of selection against gene flow. *Proceedings of the National Academy of Sciences*, 114(27):7061–7066, Jul 2017.

2. J. A. Aguirre-Liguori, B. S. Gaut, J. P. Jaramillo-Correa, M. I. Tenaillon, S. Montes-Hernández, F. García-Oliva, S. J. Hearne, and L. E. Eguiarte. 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(11):2814–2830, 2019.

3. C. Ahlmann-Eltze. *ggupset: Combination Matrix Axis for 'ggplot2' to Create 'UpSet' Plots*, 2020.

4. B. Auguie. *gridExtra: Miscellaneous Functions for "Grid" Graphics*, 2017.

5. N. Barthakur. Temperature differences between two pigmented types of corn plants. *International Journal of Biometeorology*, 18(1):70–75, Mar 1974.

6. Y. Benjamini and Y. Hochberg. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, 57(1):289–300, 1995.

7. G. Bhatia, N. Patterson, S. Sankararaman, and A. L. Price. Estimating and interpreting fst: The impact of rare variants. *Genome Research*, 23(9):1514–1521, Sep 2013.

8. A. M. Bolger, M. Lohse, and B. Usadel. Trimmomatic: a flexible trimmer for illumina sequence data. *Bioinformatics*, 30(15):2114–2120, Aug 2014.

9. K. Bomblies and J. F. Doebley. Pleiotropic effects of the duplicate maize floricaula/leafy genes zfl1 and zfl2 on traits under selection during maize domestication. *Genetics*, 172(1):519–531, Jan 2006.

10. Y. Brandvain, A. M. Kenney, L. Flagel, G. Coop, and A. L. Sweigart. Speciation and introgression between Mimulus nasutus and Mimulus guttatus. *PLoS Genetics*, 10(6):e1004410, June 2014.

11. E. S. Buckler, J. B. Holland, P. J. Bradbury, C. B. Acharya, P. J. Brown, C. Browne, E. Ersoz, S. Flint-Garcia, A. Garcia, J. C. Glaubitz, and et al. The genetic architecture of maize flowering time. *Science*, 325(5941):714–718, Aug 2009.

12. A. Canty and B. D. Ripley. *boot: Bootstrap R (S-Plus) Functions*, 2019.

13. Q. Chen, L. F. Samayoa, C. J. Yang, P. J. Bradbury, B. A. Olukolu, M. A. Neumeyer, M. C. Romay, Q. Sun, A. Lorant, E. S. Buckler, and et al. The genetic architecture of the maize progenitor, teosinte, and how it was altered during maize domestication. *PLOS Genetics*, 16(5):e1008791, May 2020.

14. R. Corbett-Detig and R. Nielsen. A Hidden Markov Model Approach for Simultaneously Estimating Local Ancestry and Admixture Time Using Next Generation Sequence Data in Samples of Arbitrary Ploidy. *PLoS Genetics*, 13(1):e1006529, Jan. 2017.

15. T. Crow, J. Ta, S. Nojoomi, M. R. Aguilar-Rangel, J. V. T. Rodríguez, D. Gates, R. Rellán-Álvarez, R. Sawers, and D. Runcie. Gene regulatory effects of a large chromosomal inversion in highland maize. *PLOS Genetics*, 16(12):e1009213, Dec 2020.

16. A. C. Davison and D. V. Hinkley. *Bootstrap Methods and Their Applications*. Cambridge University Press, Cambridge, 1997. ISBN 0-521-57391-2.

17. J. Doebley. The genetics of maize evolution. *Annual Review of Genetics*, 38(1):37–59, Nov 2004.

18. J. Doebley and A. Stec. Genetic analysis of the morphological differences between maize and teosinte. *Genetics*, 129(1):285–295, Sep 1991.

19. J. Doebley and A. Stec. Inheritance of the morphological differences between maize and teosinte: comparison of results for two f2 populations. *Genetics*, 134(2):559–570, Jun 1993.

20. J. Doebley, A. Stec, and C. Gustus. Teosinte branched1 and the origin of maize: evidence for epistasis and the evolution of dominance. *Genetics*, 141(1):333–346, Sep 1995.

21. J. Doebley, A. Stec, and L. Hubbard. The evolution of apical dominance in maize. *Nature*, 386(66246624):485–488, Apr 1997.

22. Z. Dong, O. Danilevskaya, T. Abadie, C. Messina, N. Coles, and M. Cooper. A gene regulatory network model for floral transition of the shoot apex in maize and its dynamic modeling. *PLOS ONE*, 7(8):e43450, Aug 2012.

23. Z. Dong, Y. Xiao, R. Govindarajulu, R. Feil, M. L. Siddoway, T. Nielsen, J. E. Lunn, J. Hawkins, C. Whipple, and G. Chuck. The regulatory landscape of a core maize domestication module controlling bud dormancy and growth repression. *Nature Communications*, 10(11):3810, Aug 2019.

24. J. Dorweiler, A. Stec, J. Kermicle, and J. Doebley. Teosinte glume architecture 1: A genetic locus controlling a key step in maize evolution. *Science*, 262(5131):233–235, Oct 1993.

25. C. Eberlein, M. Hénault, A. Fijarczyk, G. Charron, M. Bouvier, L. Kohn, J. Anderson, and C. Landry. Hybridization is a recurrent evolutionary stimulus in wild yeast speciation. *Nature Communications*, 10(1):923–923, Feb 2019.

26. N. B. Edelman, P. B. Frandsen, M. Miyagi, B. Clavijo, J. Davey, R. B. Dikow, G. García-Accinelli, S. M. V. Belleghem, N. Patterson, D. E. Neafsey, and et al. Genomic architecture and introgression shape a butterfly radiation. *Science*, 366(6465):594–599, Nov 2019.

27. N. C. Ellstrand, P. Meirmans, J. Rong, D. Bartsch, A. Ghosh, T. J. de Jong, P. Haccou, B.-R. Lu, A. A. Snow, C. Neal Stewart, and et al. Introgression of crop alleles into wild or weedy populations. *Annual Review of Ecology, Evolution, and Systematics*, 44(1):325–345, 2013.

28. M. C. Fontaine, J. B. Pease, A. Steele, R. M. Waterhouse, D. E. Neafsey, I. V. Sharakhov, X. Jiang, A. B. Hall, F. Catteruccia, E. Kakani, and et al. Mosquito genomics. extensive introgression in a malaria vector species complex revealed by phylogenomics. *Science (New York, N.Y.)*, 347(6217):1258524, Jan 2015.

29. K. Fukunaga, J. Hill, Y. Vigouroux, Y. Matsuoka, J. Sanchez G, K. Liu, E. S. Buckler, and J. Doebley. Genetic diversity and population structure of teosinte. *Genetics*, 169(4):2241–2254, Apr. 2005.

30. M.-A. Fustier, J.-T. Brandenburg, S. Boitard, J. Lapeyronnie, L. E. Eguiarte, Y. Vigouroux, D. Manicacci, and M. I. Tenaillon. Signatures of local adaptation in lowland and highland teosintes from whole-genome sequencing of pooled samples. *Molecular Ecology*, 26(10):2738–2756, 2017.

31. M.-A. Fustier, N. E. Martínez-Ainsworth, J. A. Aguirre-Liguori, A. Venon, H. Corti, A. Rousselet, F. Dumas, H. Dittberner, M. G. Camarena, D. Grimanelli, and et al. Common gardens in teosintes reveal the establishment of a syndrome of adaptation to altitude. *PLOS Genetics*, 15(12):e1008512, Dec 2019.

32. A. Gallavotti, Q. Zhao, J. Kyozuka, R. B. Meeley, M. K. Ritter, J. F. Doebley, M. Enrico Pè, and R. J. Schmidt. The role of barren stalk1 in the architecture of maize. *Nature*, 432(70177017):630–635, Dec 2004.

33. S. Garnier. *viridis: Default Color Maps from 'matplotlib'*, 2018.

34. E. Gonzalez-Segovia, S. Pérez-Limon, G. C. Cíntora-Martínez, A. Guerrero-Zavala, G. M. Janzen, M. B. Hufford, J. Ross-Ibarra, and R. J. H. Sawers. Characterization of introgression from the teosinte zea mays ssp. mexicana to mexican highland maize. *PeerJ*, 7:e6815, May 2019.

35. J. d. J. S. González, J. A. R. Corral, G. M. García, G. R. Ojeda, L. D. l. C. Larios, J. B. Holland, R. M. Medrano, and G. E. G. Romero. Ecogeography of teosinte. *PLOS ONE*, 13(2):e0192676, Feb 2018.

36. R. E. Green, J. Krause, A. W. Briggs, T. Maricic, U. Stenzel, M. Kircher, N. Patterson, H. Li, W. Zhai, M. H.-Y. Fritz, and et al. A draft sequence of the neandertal genome. *Science*, 328(5979):710–722, May 2010.

37. S. Haider, D. Waggott, and P. C. Boutros. *bedr: Genomic Region Processing using Tools Such as 'BEDTools', 'BEDOPS' and 'Tabix'*, 2019.

38. S. Hake and J. Ross-Ibarra. Genetic, evolutionary and plant breeding insights from the domestication of maize. *eLife*, 4:e05861, Mar 2015.

39. K. Harris and R. Nielsen. The Genetic Cost of Neanderthal Introgression. *Genetics*, 203(2):881–891, June 2016.

40. R. R. Hudson, M. Slatkin, and W. P. Maddison. Estimation of levels of gene flow from dna sequence data. *Genetics*, 132(2):583–589, Oct 1992.

41. M. B. Hufford, P. Bilinski, T. Pyhäjärvi, and J. Ross-Ibarra. Teosinte as a model system for population and ecological genomics. *Trends in Genetics*, 28(12):606–615, Dec 2012.

42. M. B. Hufford, P. Lubinksy, T. Pyhäjärvi, M. T. Devengenzo, N. C. Ellstrand, and J. Ross-Ibarra. The genomic signature of crop-wild introgression in maize. *PLoS Genetics*, 9(5):e1003477, May 2013.

43. G. M. Janzen, L. Wang, and M. B. Hufford. The extent of adaptive wild introgression in crops. *New Phytologist*, 221(3):1279–1288, 2019.

44. Y. Jiao, P. Peluso, J. Shi, T. Liang, M. C. Stitzer, B. Wang, M. S. Campbell, J. C. Stein, X. Wei, C.-S. Chin, and et al. Improved maize reference genome with single-molecule technologies. *Nature*, 546(76597659):524–527, Jun 2017.

45. I. Juric, S. Aeschbacher, and G. Coop. The Strength of Selection against Neanderthal Introgression. *PLoS Genetics*, 12(11):e1006340, Nov. 2016.

46. A. M. Kenney and A. L. Sweigart. Reproductive isolation and introgression between sympatric mimulus species. *Molecular Ecology*, 25(11):2499–2517, Jun 2016.

47. T. S. Korneliussen, A. Albrechtsen, and R. Nielsen. Angsd: Analysis of next generation sequencing data. *BMC bioinformatics*, 15(1):356, Dec 2014.

48. J. Köster and S. Rahmann. Snakemake—a scalable bioinformatics workflow engine. *Bioinformatics*, 28(19):2520–2522, Oct 2012.

49. N. Lauter, C. Gustus, A. Westerbergh, and J. Doebley. The inheritance and evolution of leaf pigmentation and pubescence in teosinte. *Genetics*, 167(4):1949–1959, Aug 2004.

50. V. Le Corre, M. Siol, Y. Vigouroux, M. I. Tenaillon, and C. Délye. Adaptive introgression from maize has facilitated the establishment of teosinte as a noxious weed in europe. *Proceedings of the National Academy of Sciences*, 117(41):25618–25627, Oct 2020.

51. H. Li. Improving snp discovery by base alignment quality. *Bioinformatics*, 27(8):1157–1158, Apr 2011.

52. H. Li. Aligning sequence reads, clone sequences and assembly contigs with bwa-mem. *arXiv [q-bio]*, May 2013. arXiv: 1303.3997.

53. H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin, and . G. P. D. P. Subgroup. The sequence alignment/map format and samtools. *Bioinformatics*, 25(16):2078–2079, Aug 2009.

54. Y.-x. Li, C. Li, P. J. Bradbury, X. Liu, F. Lu, C. M. Romay, J. C. Glaubitz, X. Wu, B. Peng, Y. Shi, and et al. Identification of genetic variants associated with maize flowering time using an extremely large multi-genetic background population. *The Plant Journal*, 86(5):391–402, 2016.

55. Z. Lin, X. Li, L. M. Shannon, C.-T. Yeh, M. L. Wang, G. Bai, Z. Peng, J. Li, H. N. Trick, T. E. Clemente, and et al. Parallel domestication of the shattering1 genes in cereals. *Nature Genetics*, 44(6):720–724, May 2012.

56. P.-R. Loh, M. Lipson, N. Patterson, P. Moorjani, J. K. Pickrell, D. Reich, and B. Berger. Inferring admixture histories of human populations using linkage disequilibrium. *Genetics*, 193(4):1233–1254, Apr 2013.

57. J. Mammadov, R. Buyyarapu, S. K. Guttikonda, K. Parliament, I. Y. Abdurakhmonov, and S. P. Kumpatla. Wild relatives of maize, rice, cotton, and soybean: Treasure troves for tolerance to biotic and abiotic stresses. *Frontiers in Plant Science*, 9, 2018.

58. C. H. Martin, J. S. Cutler, J. P. Friel, C. Dening Touokong, G. Coop, and P. C. Wainwright. Complex histories of repeated gene flow in cameroon crater lake cichlids cast doubt on one of the clearest examples of sympatric speciation. *Evolution*, 69(6):1406–1422, Jun 2015.

59. S. H. Martin, J. W. Davey, C. Salazar, and C. D. Jiggins. Recombination rate variation shapes barriers to introgression across butterfly genomes. *PLOS Biology*, 17(2):e2006288, Feb 2019.

60. J. Meisner and A. Albrechtsen. Inferring population structure and admixture proportions in low-depth ngs data. *Genetics*, 210(2):719–731, Oct 2018.

61. R. S. Meyer and M. D. Purugganan. Evolution of crop species: genetics of domestication and diversification. *Nature Reviews Genetics*, 14(1212):840–852, Dec 2013.

62. P. Moorjani, N. Patterson, J. N. Hirschhorn, A. Keinan, L. Hao, G. Atzmon, E. Burns, H. Ostrer, A. L. Price, and D. Reich. The history of african gene flow into southern europeans, levantines, and jews. *PLOS Genetics*, 7(4):e1001373, Apr 2011.

63. S. P. Moose, N. Lauter, and S. R. Carlson. The maize macrohairless1 locus specifically promotes leaf blade macrohair initiation and responds to factors regulating leaf identity. *Genetics*, 166(3):1451–1461, Mar 2004.

64. G. Moya-Raygoza. Early development of leaf trichomes is associated with decreased damage in teosinte, compared with maize, by spodoptera frugiperda (lepidoptera: Noctuidae). *Annals of the Entomological Society of America*, 109(5):737–743, Sep 2016.

65. T. C. Nelson, A. M. Stathos, D. D. Vanderpool, F. R. Finseth, Y.-w. Yuan, and L. Fishman. Ancient and recent introgression shape the evolutionary history of pollinator adaptation and speciation in a model monkeyflower radiation (mimulus section erythranthe). *PLOS Genetics*, 17(2):e1009095, Feb 2021.

66. F. Ogut, Y. Bian, P. J. Bradbury, and J. B. Holland. Joint-multiple family linkage analysis predicts within-family variation better than single-family analysis of the maize nested association mapping population. *Heredity*, 114(6):552–563, Jan. 2015.

67. A. M. O'Brien, R. J. H. Sawers, S. Y. Strauss, and J. Ross-Ibarra. Adaptive phenotypic divergence in an annual grass differs across biotic contexts. *Evolution*, 73(11):2230–2246, 2019.

68. J. B. Pease, D. C. Haak, M. W. Hahn, and L. C. Moyle. Phylogenomics reveals three sources of adaptive variation during a rapid radiation. *PLOS Biology*, 14(2):e1002379, Feb 2016.

69. T. L. Pedersen. *tidygraph: A Tidy API for Graph Manipulation*, 2020.

70. B. M. Peter. Admixture, population structure, and f-statistics. *Genetics*, 202(4):1485–1501, Apr 2016.

71. K. S. Pfennig, A. L. Kelly, and A. A. Pierce. Hybridization as a facilitator of species range expansion. *Proceedings of the Royal Society B: Biological Sciences*, 283(1839):20161329, Sep 2016.

72. H. Piazena. The effect of altitude upon the solar uv-b and uv-a irradiance in the tropical chilean andes. *Solar Energy*, 57(2):133–140, Aug 1996.

73. D. R. Piperno and K. V. Flannery. The earliest archaeological maize (zea mays l.) from highland mexico: New accelerator mass spectrometry dates and their implications. *Proceedings of the National Academy of Sciences*, 98(4):2101–2103, Feb 2001.

74. D. R. Piperno, A. J. Ranere, I. Holst, J. Iriarte, and R. Dickau. Starch grain and phytolith evidence for early ninth millennium b.p. maize from the central balsas river valley, mexico. *Proceedings of the National Academy of Sciences*, 106(13):5019–5024, Mar 2009.

75. I. Portwood, John L, M. R. Woodhouse, E. K. Cannon, J. M. Gardiner, L. C. Harper, M. L. Schaeffer, J. R. Walsh, T. Z. Sen, K. T. Cho, D. A. Schott, and et al. Maizegdb 2018: the maize multi-genome genetics and genomics database. *Nucleic Acids Research*, 47(D1):D1146–D1154, Jan 2019.

76. D. L. Powell, M. García-Olazábal, M. Keegan, P. Reilly, K. Du, A. P. Díaz-Loyo, S. Banerjee, D. Blakkan, D. Reich, P. Andolfatto, and et al. Natural hybridization reveals incompatible alleles that cause melanoma in swordtail fish. *Science*, 368(6492):731–736, May 2020.

77. D. C. Presgraves. The molecular evolutionary basis of species formation. *Nature Reviews Genetics*, 11(33):175–180, Mar 2010.

78. M. D. Purugganan. Evolutionary insights into the nature of plant domestication. *Current Biology*, 29(14):R705–R714, Jul 2019.

79. T. Pyhäjärvi, M. B. Hufford, S. Mezmouk, and J. Ross-Ibarra. Complex patterns of local adaptation in teosinte. *Genome Biology and Evolution*, 5(9):1594–1609, 2013.

80. A. R. Quinlan and I. M. Hall. Bedtools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, 26(6):841–842, Mar 2010.

81. R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2019.

82. L. H. Rieseberg, O. Raymond, D. M. Rosenthal, Z. Lai, K. Livingstone, T. Nakazato, J. L. Durphy, A. E. Schwarzbach, L. A. Donovan, and C. Lexer. Major ecological transitions in wild sunflowers facilitated by hybridization. *Science*, 301(5637):1211–1216, Aug 2003.

83. J. G. Rodríguez, G. Sánchez, B. M. Baltazar, L. L. D. l. Cruz, F. Santacruz-Ruvalcaba, P. J. Ron, and J. B. Schoper. Characterization of floral morphology and synchrony among zea species in mexico. *Maydica*, pages 383–398, 2006.

84. F. Rodríguez-Zapata, A. C. Barnes, K. A. Blöcher-Juárez, D. Gates, A. Kur, L. Wang, G. M. Janzen, S. Jensen, J. M. Estévez-Palmas, T. Crow, and et al. Teosinte introgression modulates phosphatidylcholine levels and induces early maize flowering time. *bioRxiv*, Feb 2021.

85. J. Ross-Ibarra, M. Tenaillon, and B. S. Gaut. Historical divergence and gene flow in the genus zea. *Genetics*, 181(4):1399–1413, Apr 2009.

86. B. A. Rowan, D. Heavens, T. R. Feuerborn, A. J. Tock, I. R. Henderson, and D. Weigel. An ultra high-density arabidopsis thaliana crossover map that refines the influences of structural variation and epigenetic features. *Genetics*, 213(3):771–787, Nov 2019.

87. S. Sankararaman, S. Mallick, M. Dannemann, K. Prüfer, J. Kelso, S. Pääbo, N. Patterson, and D. Reich. The genomic landscape of neanderthal ancestry in present-day humans. *Nature*, 507(74927492):354–357, Mar 2014.

88. P. S. Schnable, D. Ware, R. S. Fulton, J. C. Stein, F. Wei, S. Pasternak, C. Liang, J. Zhang, L. Fulton, T. A. Graves, and et al. The b73 maize genome: Complexity, diversity, and dynamics. *Science*, 326(5956):1112–1115, Nov 2009.

89. M. Schumer, C. Xu, D. L. Powell, A. Durvasula, L. Skov, C. Holland, J. C. Blazier, S. Sankararaman, P. Andolfatto, G. G. Rosenthal, and M. Przeworski. Natural selection interacts with recombination to shape the evolution of hybrid genomes. *Science*, 360(6389):656–660, May 2018.

90. B. Sigmon and E. Vollbrecht. Evidence of selection at the ramosa1 locus during maize domestication. *Molecular Ecology*, 19(7):1296–1311, 2010.

91. L. Skotte, T. S. Korneliussen, and A. Albrechtsen. Estimating individual admixture proportions from next generation sequencing data. *Genetics*, 195(3):693–702, Nov 2013.

92. S. Soraggi, C. Wiuf, and A. Albrechtsen. Powerful inference with the d-statistic on low-coverage whole-genome data. *G3: Genes—Genomes—Genetics*, 8(2):551–566, Feb 2018.

93. D. Sosso, D. Luo, Q.-B. Li, J. Sasse, J. Yang, G. Gendrot, M. Suzuki, K. E. Koch, D. R. McCarty, P. S. Chourey, and et al. Seed filling in domesticated maize and rice depends on sweet-mediated hexose transport. *Nature Genetics*, 47(1212):1489–1493, Dec 2015.

94. J. E. Staller. High altitude maize (zea mays l.) cultivation and endemism in the lake titicaca basin. *Journal of Botany Research*, 1(11), Nov 2016.

95. A. E. Stapleton and V. Walbot. Flavonoids can protect maize dna from the induction of ultraviolet radiation damage. *Plant Physiology*, 105(3):881–889, Jul 1994.

96. M. C. Stitzer and J. Ross-Ibarra. Maize domestication and gene interaction. *New Phytologist*, 220(2):395–408, 2018.

97. A. Suvorov, B. Y. Kim, J. Wang, E. E. Armstrong, D. Peede, E. R. R. D'Agostino, D. K. Price, P. Wadell, M. Lang, J. R. David, and et al. Widespread introgression across a phylogeny of 155 drosophila genomes. *bioRxiv*, page 33, Jan 2021.

98. M. I. Tenaillon, M. C. Sawkins, A. D. Long, R. L. Gaut, J. F. Doebley, and B. S. Gaut. Patterns of dna sequence polymorphism along chromosome 1 of maize (zea mays ssp. mays l.). *Proceedings of the National Academy of Sciences*, 98(16):9161–9166, Jul 2001.

99. L. Trimble, S. Shuler, and W. F. Tracy. Characterization of five naturally occurring alleles at the sugary1 locus for seed composition, seedling emergence, and isoamylase1 activity. *Crop Science*, 56(4):1927–1939, 2016.

100. S. Tusso, B. P. S. Nieuwenhuis, F. J. Sedlazeck, J. W. Davey, D. C. Jeffares, and J. B. W. Wolf. Ancestral admixture is the main determinant of global biodiversity in fission yeast. *Molecular Biology and Evolution*, 36(9):1975–1989, Sep 2019.

101. J. van Heerwaarden, J. Ross-Ibarra, J. Doebley, J. C. Glaubitz, J. d. J. S. Gonzalez, B. S. Gaut, and L. E. Eguiarte. Fine scale genetic structure in the wild ancestor of maize (zea mays ssp. parviglumis). *Molecular ecology*, 2010.

102. C. Veller, N. B. Edelman, P. Muralidhar, and M. A. Nowak. Recombination, variance in genetic relatedness, and selection against introgressed DNA. *bioRxiv*, page 846147, Nov. 2019.

103. E. Vollbrecht, P. S. Springer, L. Goh, E. S. Buckler Iv, and R. Martienssen. Architecture of floral branch systems in maize and related grasses. *Nature*, 436(70547054):1119–1126, Aug 2005.

104. H. Wang, T. Nussbaum-Wagler, B. Li, Q. Zhao, Y. Vigouroux, M. Faller, K. Bomblies, L. Lukens, and J. F. Doebley. The origin of the naked grains of maize. *Nature*, 436(70517051):714–719, Aug 2005.

105. L. Wang, T. M. Beissinger, A. Lorant, C. Ross-Ibarra, J. Ross-Ibarra, and M. B. Hufford. The interplay of demography and selection during maize domestication and expansion. *Genome biology*, 18(1):215, Nov. 2017.

106. Z. Wang, T. Ueda, and J. Messing. Characterization of the maize prolamin box-binding factor-1 (pbf-1) and its role in the developmental regulation of the zein multigene family. *Gene*, 223(1):321–332, Nov 1998.

107. S. R. Whitt, L. M. Wilson, M. I. Tenaillon, B. S. Gaut, and E. S. Buckler. Genetic diversity and selection in the maize starch pathway. *Proceedings of the National Academy of Sciences*, 99(20):12959–12962, Oct 2002.

108. H. Wickham, M. Averick, J. Bryan, W. Chang, L. D. McGowan, R. François, G. Grolemund, A. Hayes, L. Henry, J. Hester, M. Kuhn, T. L. Pedersen, E. Miller, S. M. Bache, K. Müller, J. Ooms, D. Robinson, D. P. Seidel, V. Spinu, K. Takahashi, D. Vaughan, C. Wilke, K. Woo, and H. Yutani. Welcome to the tidyverse. *Journal of Open Source Software*, 4(43):1686, 2019.

109. H. G. Wilkes. *Teosinte: The Closest Relative of Maize*. The Bussey Institute of Harvard University, 1967.

110. H. G. Wilkes. Hybridization of maize and teosinte, in Mexico and Guatemala and improvement of maize. *Economic Botany*, 31(3):254–293, 1977.

111. D. M. Wills, Z. Fang, A. M. York, J. B. Holland, and J. F. Doebley. Defining the role of the mads-box gene, zea agamous-like1, a target of selection during maize domestication. *Journal of Heredity*, 109(3):333–338, Mar 2018.

112. D. M. Wills, C. J. Whipple, S. Takuno, L. E. Kursel, L. M. Shannon, J. Ross-Ibarra, and J. F. Doebley. From many, one: Genetic control of prolificacy during maize domestication. *PLoS Genetics*, 9(6), Jun 2013.

113. S. W. Wingett and S. Andrews. Fastq screen: A tool for multi-genome mapping and quality control. *F1000Research*, 7:1338, Sep 2018.

114. C. J. Yang, L. F. Samayoa, P. J. Bradbury, B. A. Olukolu, W. Xue, A. M. York, M. R. Tuholski, W. Wang, L. L. Daskalska, M. A. Neumeyer, and et al. The genetic architecture of teosinte catalyzed and constrained maize domestication. *Proceedings of the National Academy of Sciences*, 116(12):5643–5652, Mar 2019.

115. M. P. Zuellig and A. L. Sweigart. Gene duplicates cause hybrid lethality between sympatric species of mimulus. *PLOS Genetics*, 14(4):e1007130, Apr 2018.

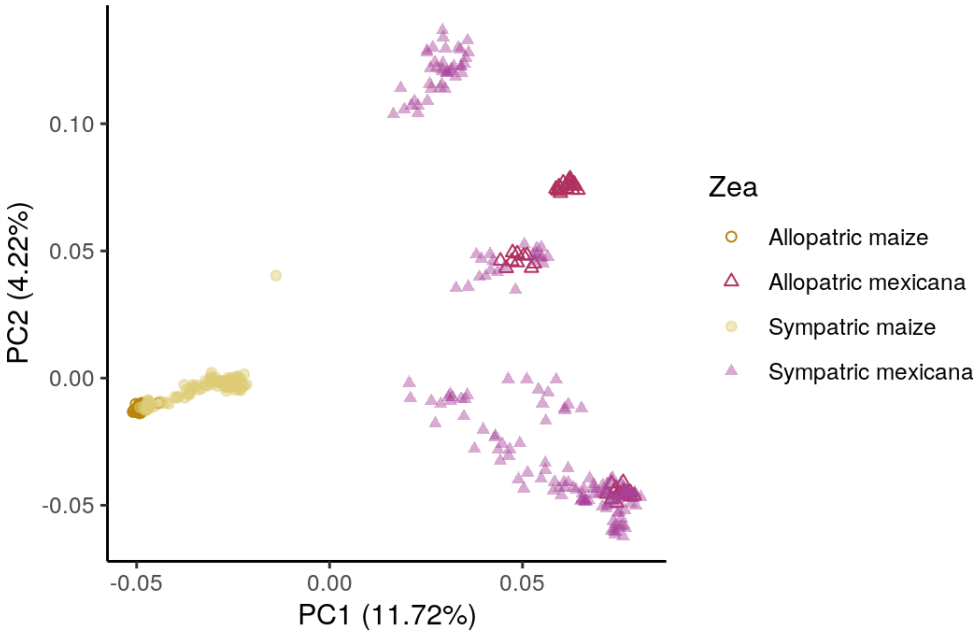

Fig S2. PCA. First and second principal components from the genomewide genetic covariance matrix relating sympatric and allopatric maize and *mexicana* individuals (PCAngsd). PC1 separates maize and *mexicana* subspecies while PC2 differentiates genetic clusters within *mexicana*.

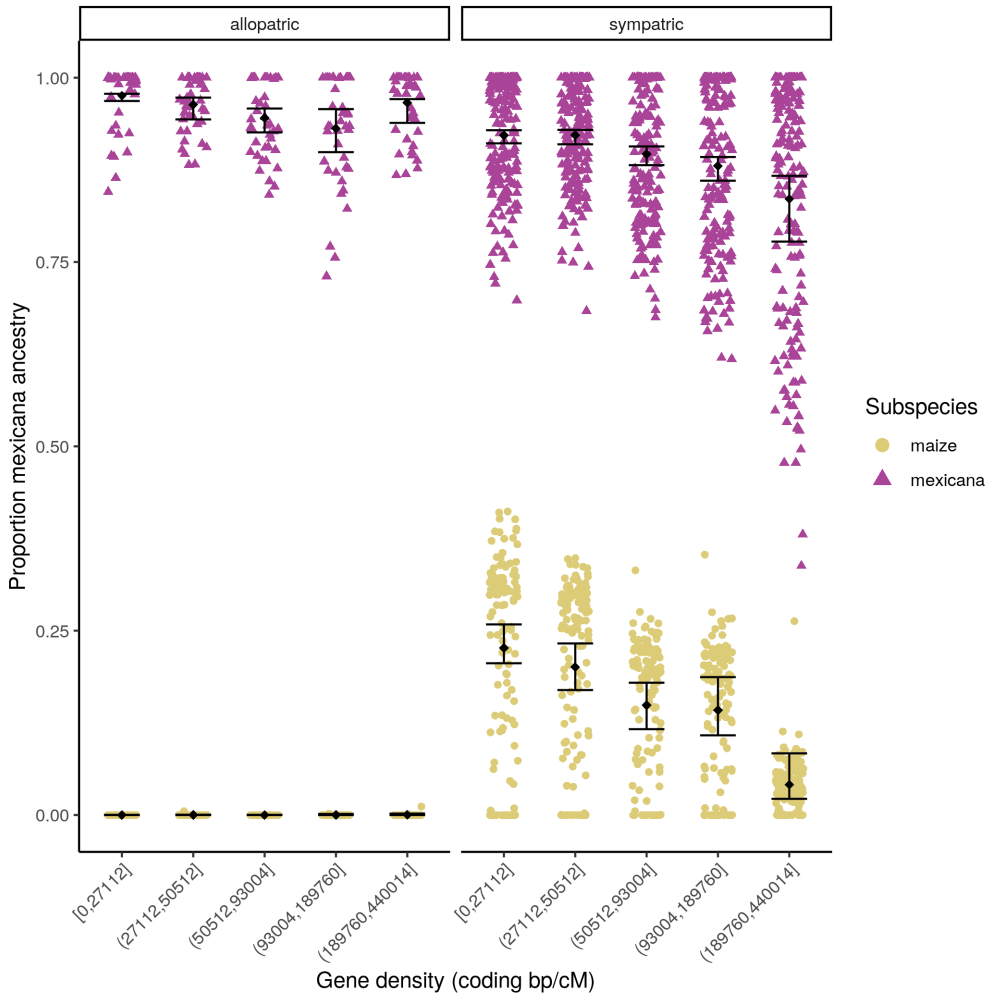

Fig S3. *Mexicana* ancestry by coding bp per cM. Inferred *mexicana* ancestry in allopatric reference populations (left) and sympatric maize and *mexicana* populations (right) using NGSAdmix (K=2) by coding density quintiles. Group mean and 95% percentile bootstrap confidence interval (n = 100) are depicted in black. Ancestry estimates for each individual are shown as points, colored by *Zea* subspecies, and points are jittered for better visualization.

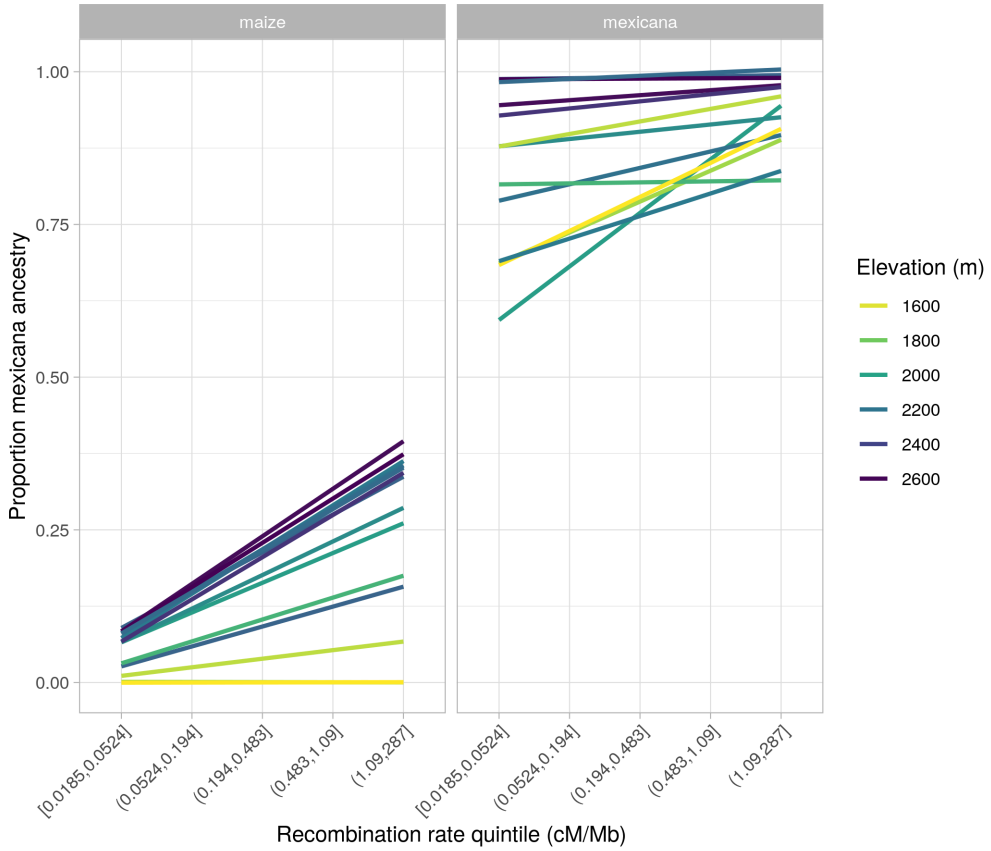

Fig S4. *Mexicana* ancestry across recombination quintiles by elevation. Estimated linear relationship between proportion *mexicana* ancestry (NGSAdmix) and recombination rate quintile for each sympatric population, colored by elevation. Linear models were fit using *lm()* in R on individuals' ancestry estimates per quintile.

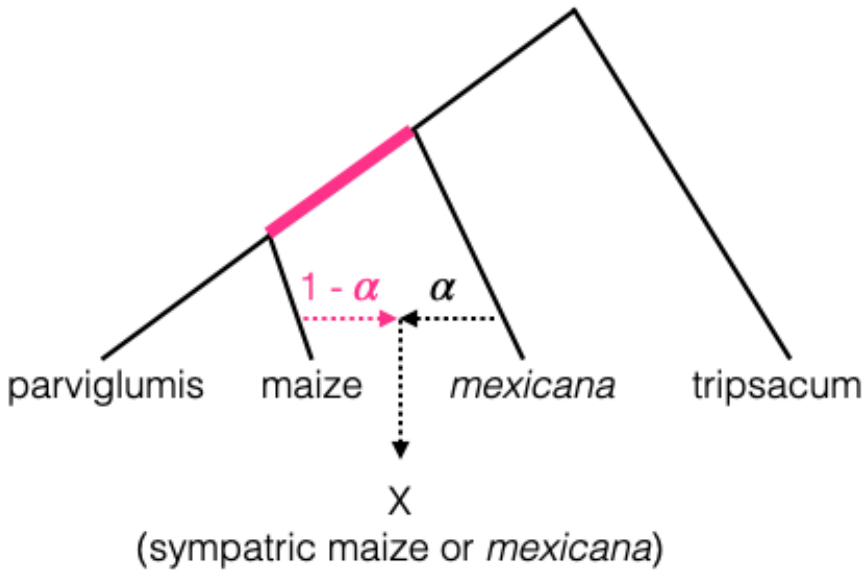

Fig S5. Population tree Phylogenetic tree assumed when estimating the ratio of f_4 statistics. The pink branch represents the shared drift between maize and parviglumis that is introduced to the allopatric population via admixture of proportion $1 - \alpha$. We used only plants from the Amecameca site in our allopatric *mexicana* reference group for this analysis because that site showed no evidence of previous admixture.

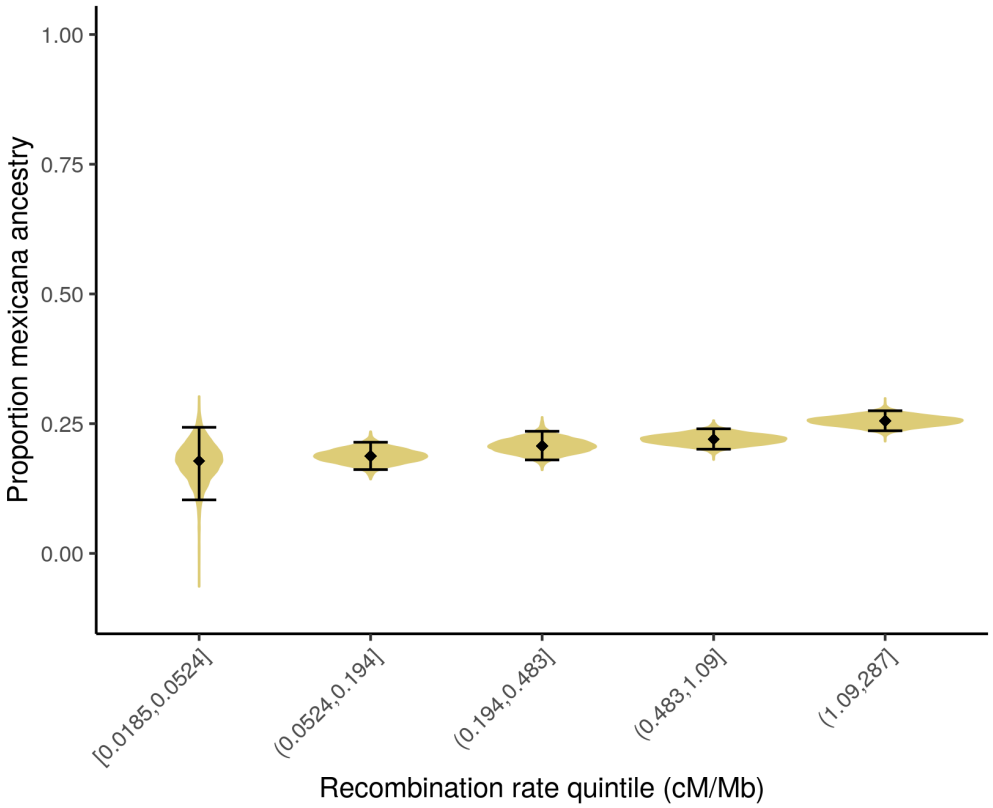

Fig S6. f_4 **ancestry in maize by recombination rate**. Estimated *mexicana* ancestry in sympatric maize landrace samples using f_4 ratio. Mean ancestry per recombination rate quintile and 95% percentile bootstrap confidence interval (n = 10,000) are depicted in black. Violin plots show the density of ancestry estimates for individual bootstraps re-sampled within quintiles.

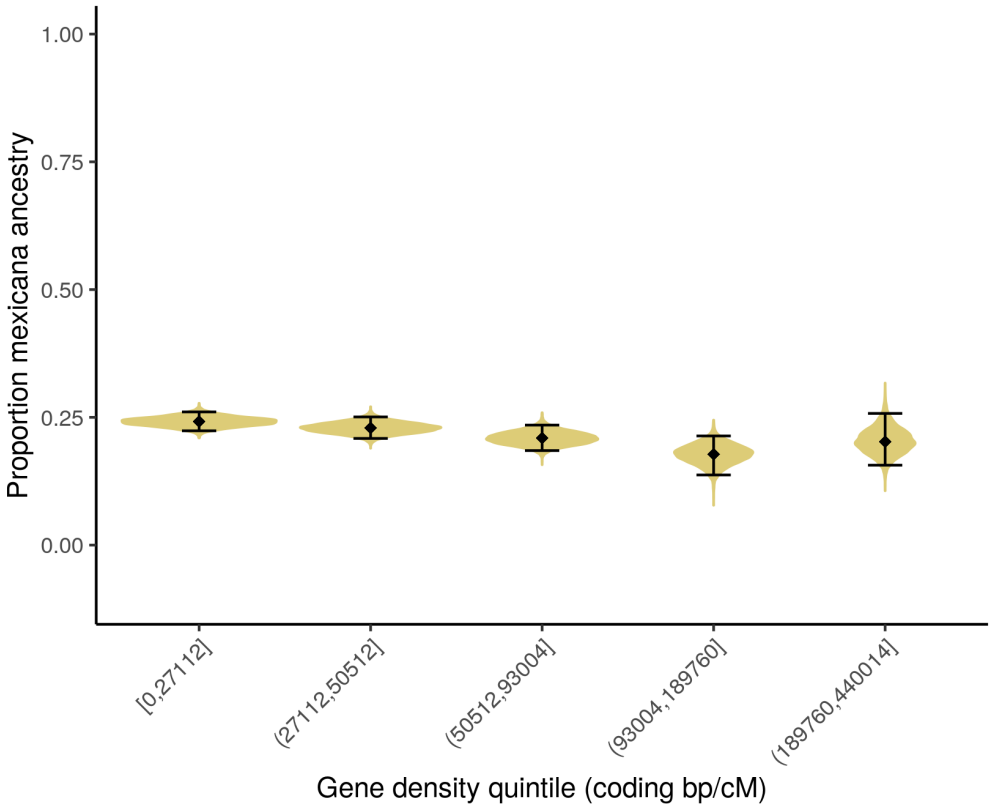

Fig S7. f_4 **ancestry in maize by coding bp per cM**. Estimated *mexicana* ancestry in sympatric maize landrace samples using f_4 ratio. Mean ancestry for each coding bp/cM quintile and 95% percentile bootstrap confidence interval (n = 10,000) are depicted in black. Violin plots show the density of ancestry estimates for individual bootstraps re-sampled within quintiles.

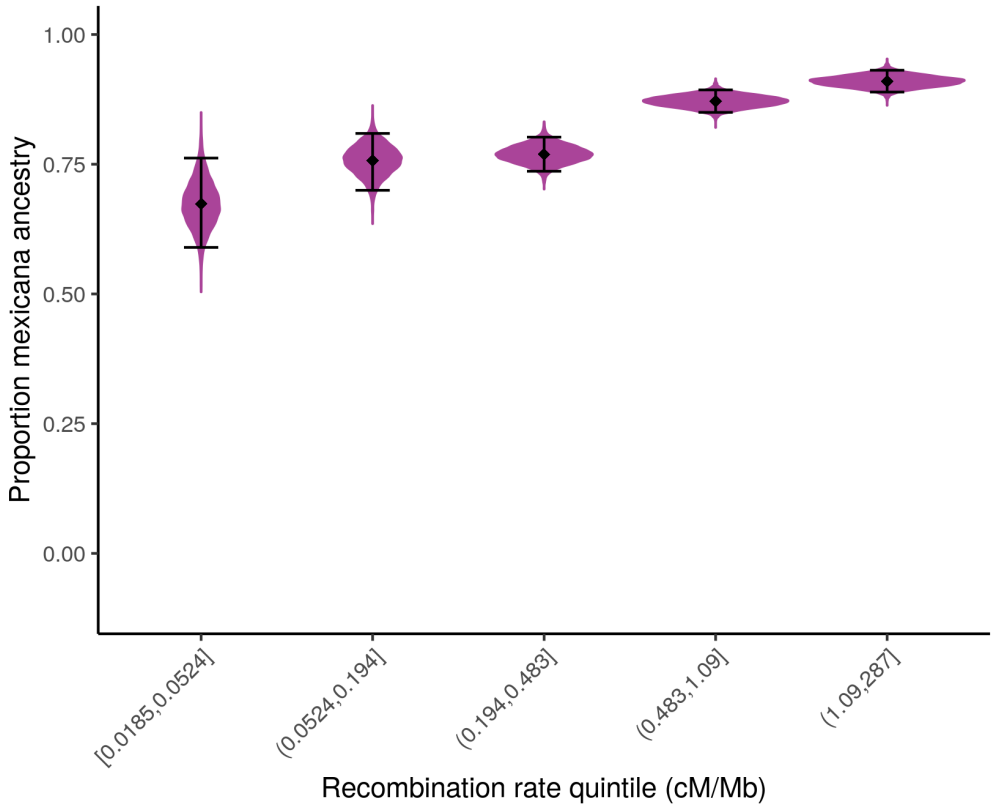

Fig S8. f_4 **ancestry in maize by recombination rate**. Estimated *mexicana* ancestry in sympatric mexicana samples using f_4 ratio. Mean ancestry per recombination rate quintile and 95% percentile bootstrap confidence interval (n = 10,000) are depicted in black. Violin plots show the density of ancestry estimates for individual bootstraps re-sampled within quintiles.

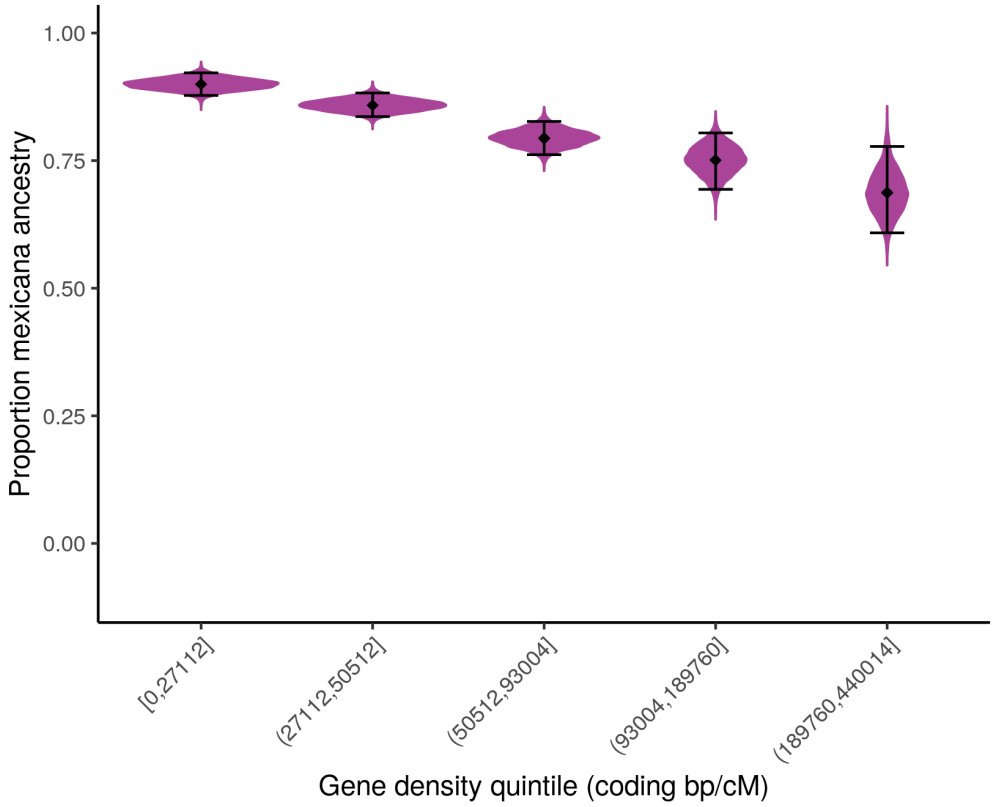

Fig S9. f_4 ancestry in mexicana by coding bp per cM. Estimated *mexicana* ancestry in sympatric mexicana samples using f_4 ratio. Mean ancestry for each coding bp/cM quintile and 95% percentile bootstrap confidence interval (n = 10,000) are depicted in black. Violin plots show the density of ancestry estimates for individual bootstraps re-sampled within quintiles.

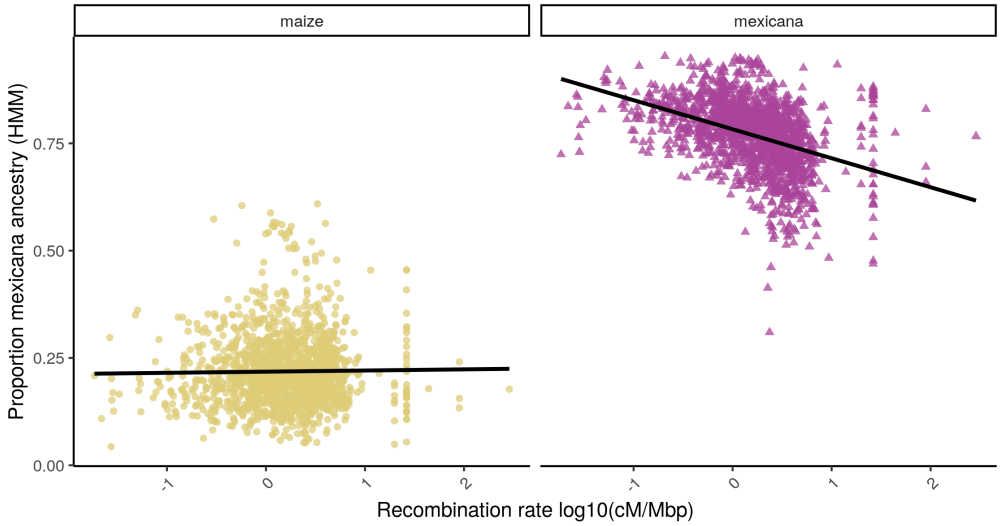

Fig S10. Local *mexicana* ancestry in 1 cM windows by recombination rate. Estimated *mexicana* ancestry in sympatric maize and *mexicana* samples using ancestry_hmm. Each point is a 1 cM genomic window and the line shows the best linear model fit for mean *mexicana* ancestry by recombination rate on a log scale.

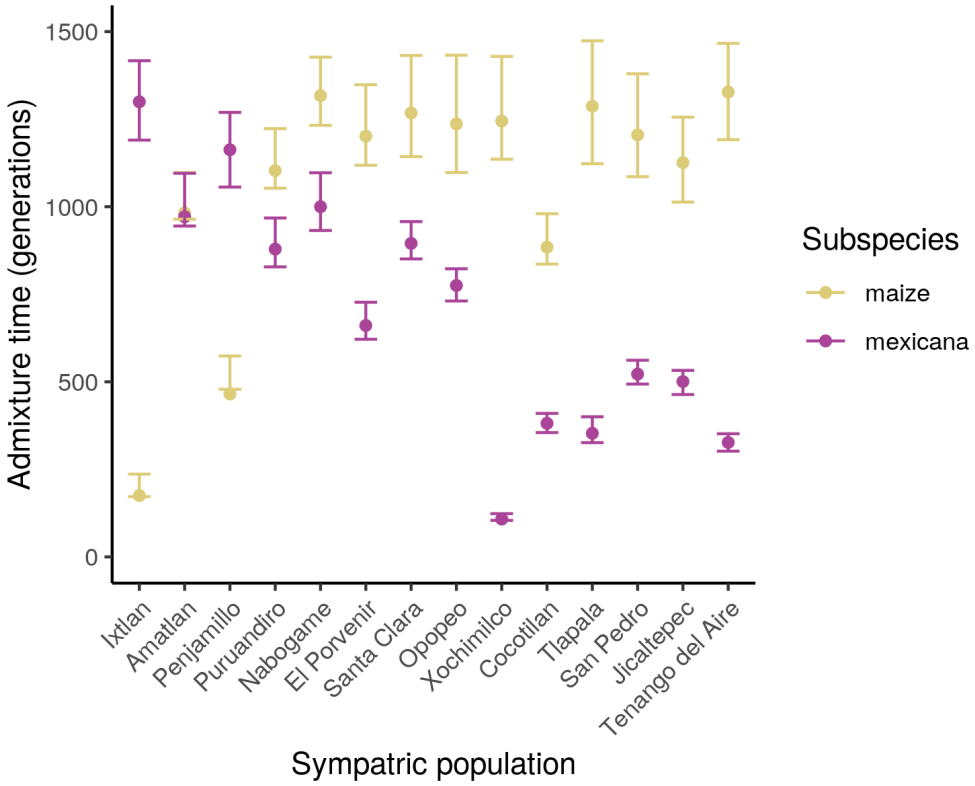

Fig S11. Time since admixture. Estimated generations since admixture under a single-pulse model for each sympatric maize and *mexicana* population, with 95% percentile confidence intervals based on 100 bootstrap samples of genomic blocks (1,000 SNPs per block). Estimates and bootstraps were produced during ancestry_hmm model fitting for local ancestry inference. Populations are ordered left to right by increasing elevation.

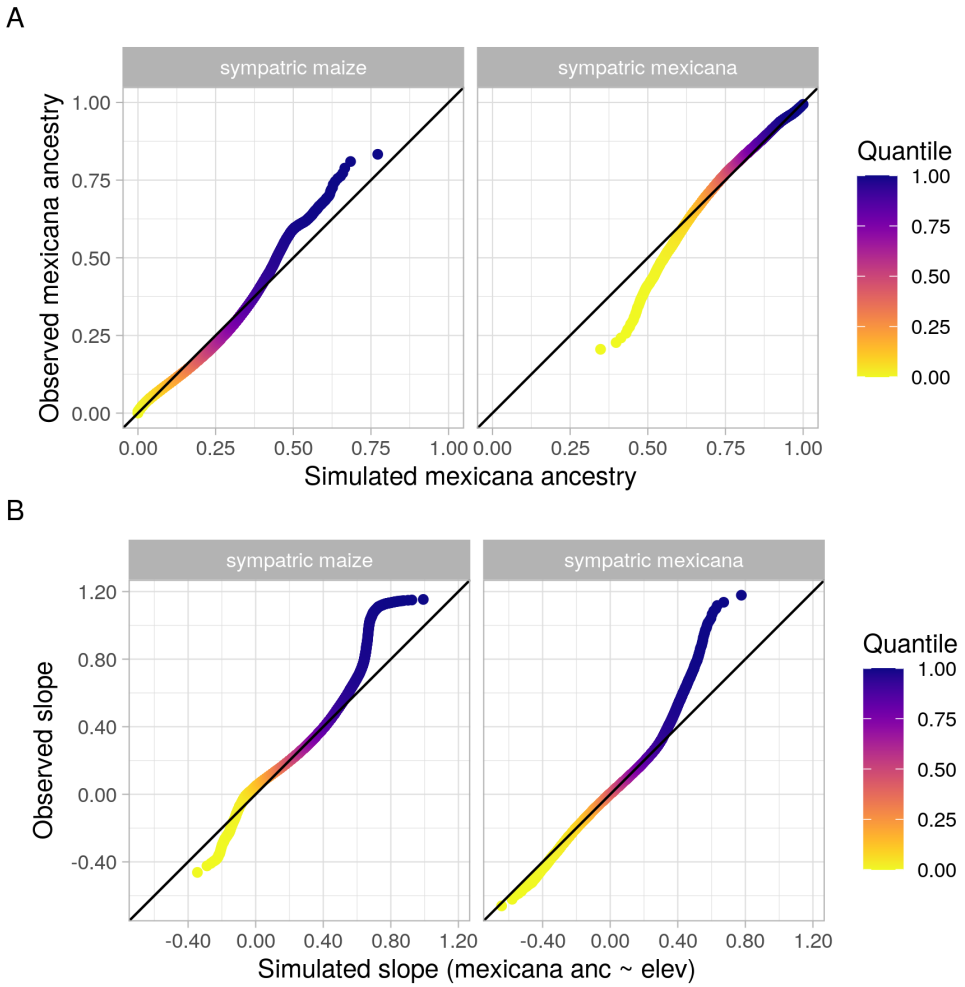

Fig S12. Quantile comparison of observed data vs. MVN normal null model (A) QQ-plot of simulated vs. observed mean ancestry at individual loci across all sympatric populations. (B) QQ-plot of simulated vs. observed slopes from the linear model *mexicana* ancestry ∼ elevation at individual loci.

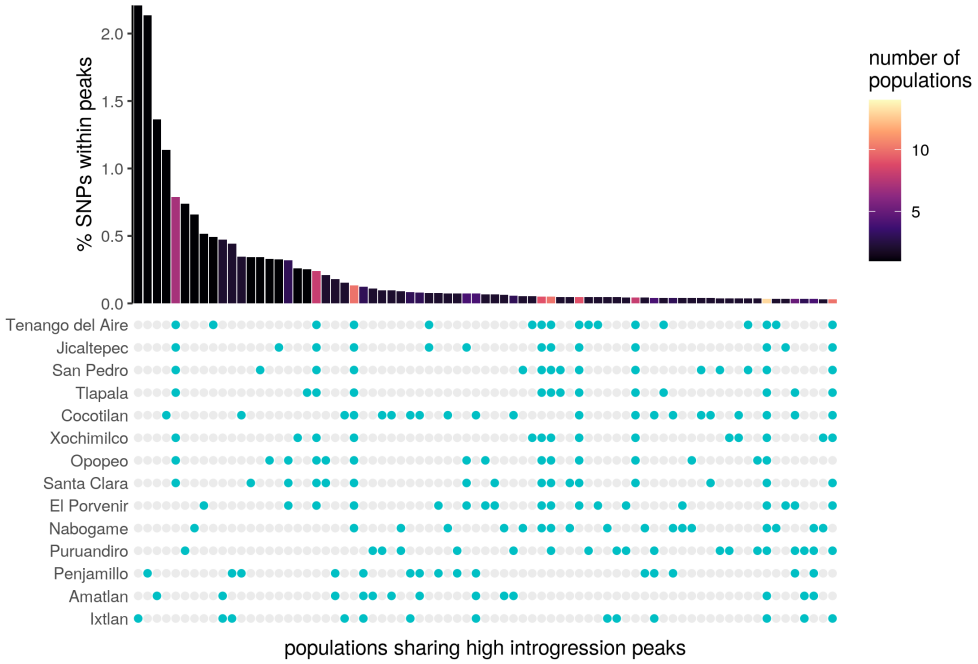

Fig S13. High introgression peaks shared across sympatric maize populations Here we show the 75 most common combinations of populations that share ancestry peaks (introgressed ancestry > 2 s.d. above each population's mean ancestry). Bar height represents the percent of SNPs genomewide within peaks shared by the populations highlighted in blue below. Populations are ordered from high (top) to low elevation. See S14 for sympatric *mexicana* equivalent visualization.

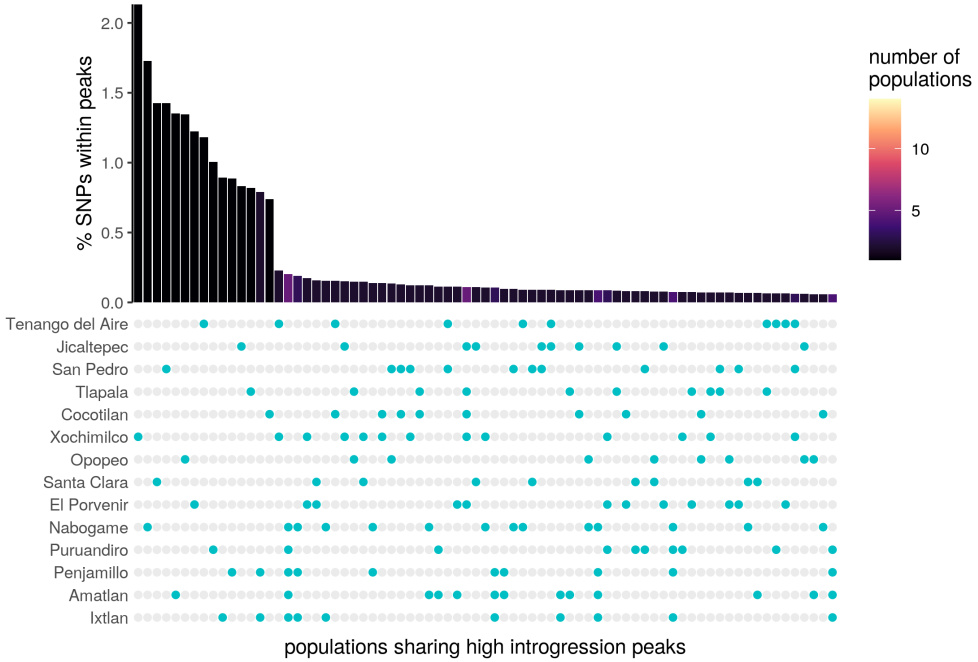

Fig S14. High introgression peaks shared across sympatric *mexicana* populations Here we show the 75 most common combinations of populations that share ancestry peaks (introgressed ancestry > 2 s.d. above each population's mean ancestry). Bar height represents the percent of SNPs genomewide within peaks shared by the populations highlighted in blue below. Populations are ordered from high (top) to low elevation.

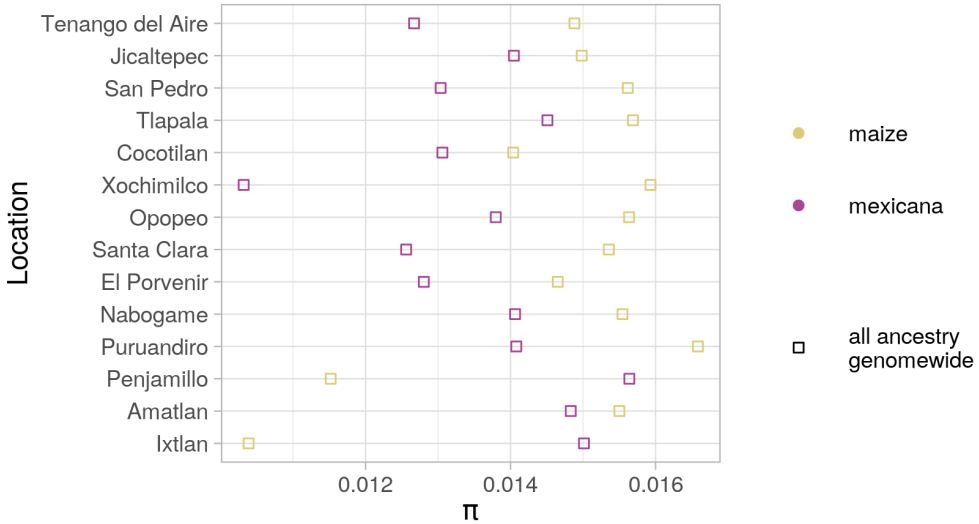

Fig S15. Diversity (π) within maize ancestry Each point summarises pairwise genetic diversity (π) for regions genomewide with high-confidence homozygous *mexicana* ancestry, calculated separately for the maize and *mexicana* populations at each sampled location.

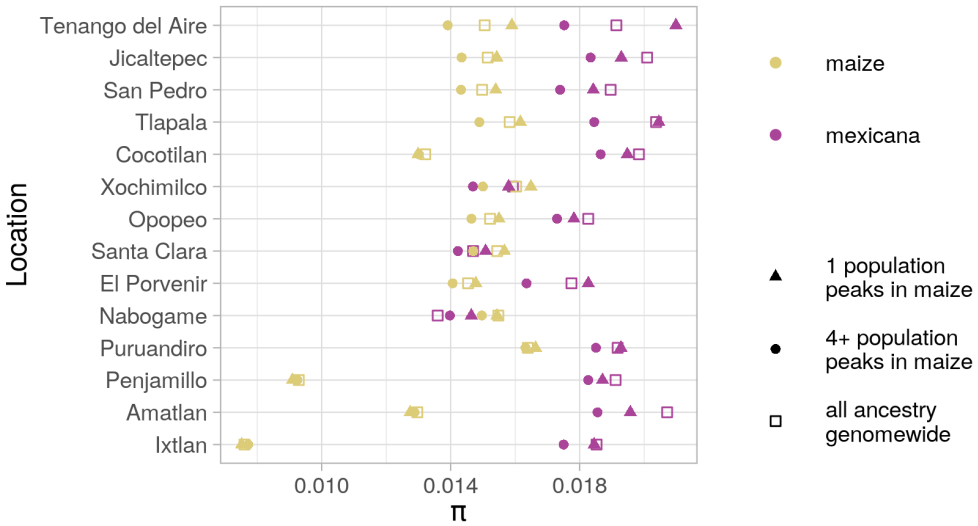

Fig S16. Diversity (π) within *mexicana* ancestry Each point summarises pairwise genetic diversity (π) for genomic regions with high-confidence homozygous *mexicana* ancestry, calculated separately for the maize and *mexicana* populations at each sampled location. Within-*mexicana* ancestry π is calculated and plotted separately for three subsets of the genome: introgression peaks ($>$ 2 s.d. above the mean) found in the focal maize population only, peaks shared between the focal maize and at least 3 other maize populations, and a genomewide estimate.

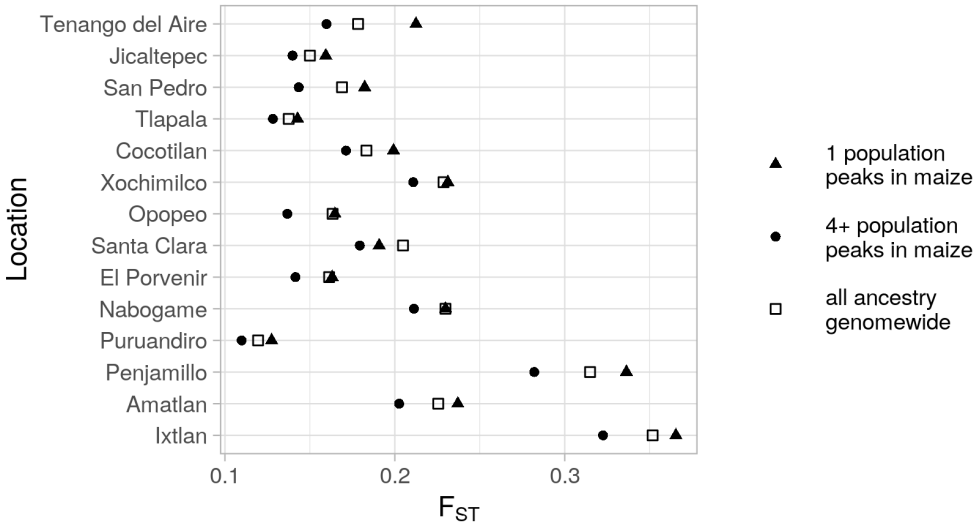

Fig S17. Differentiation (F_{ST}) between introgressed ancestry tracts and local *mexicana* Each point summarises F_{ST} between *mexicana* ancestry tracts within a focal maize population and *mexicana* ancestry tracts within the local *mexicana* population sampled at the same site. Within-*mexicana* ancestry F_{ST} is presented separately for three subsets of the genome: introgression peaks found in the focal maize population only, peaks shared between the focal maize and at least 3 other maize populations, and a genomewide estimate.

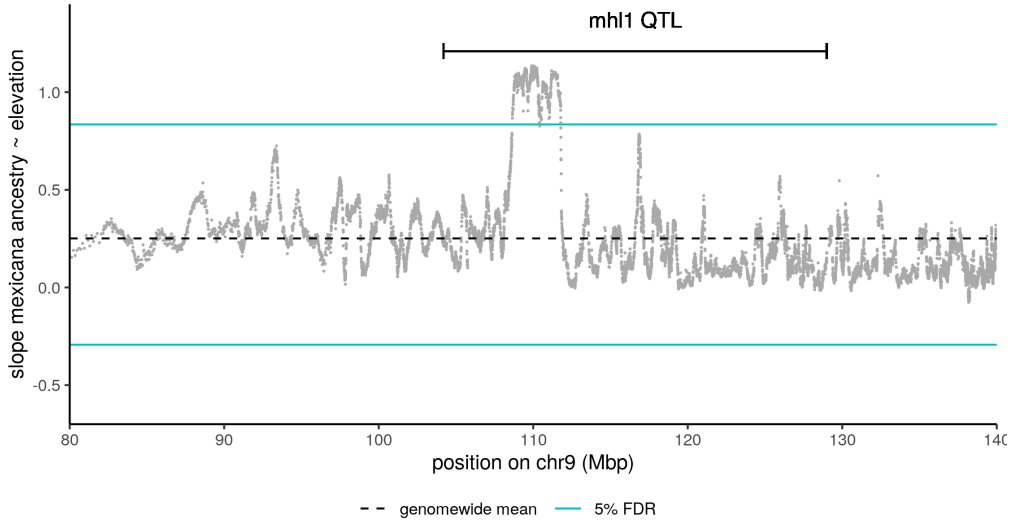

Fig S18. Ancestry slope with elevation at mhl1 locus. Slope of introgressed *mexicana* ancestry proportion in sympatric maize over a 1 km gain in elevation, zoomed in on the mhl1 QTL region on chromosome 9. Coordinates for the 3 Mb outlier region within this QTL are 9:108640415-111788150.

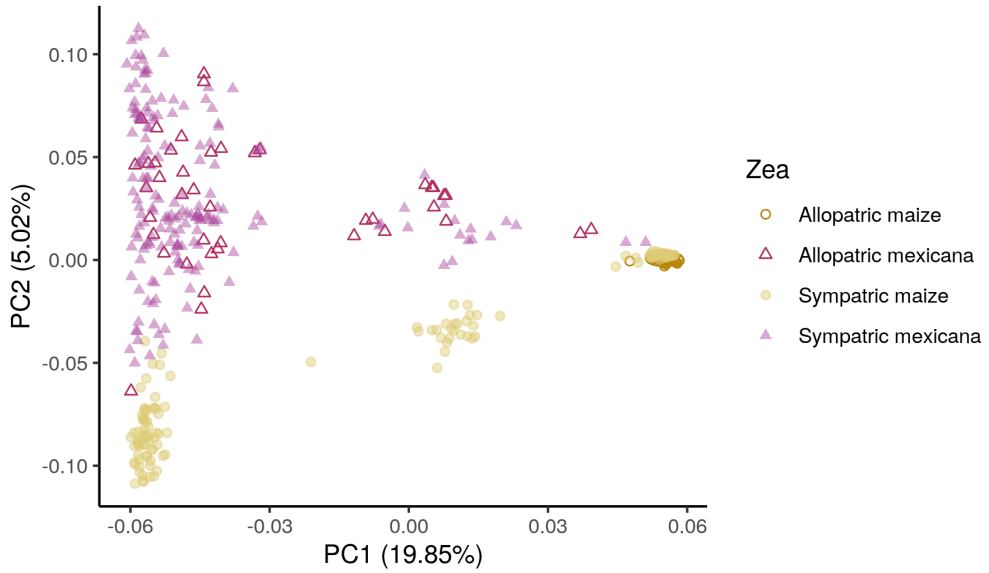

Fig S19. PCA of putative mhl1 inversion. Principal components analysis of all SNPs in the 3 Mb outlier region within the mhl1 QTL region that shows a steep increase in introgressed *mexicana* ancestry across elevation (>5% FDR). This region on chromosome 9 is a putative inversion (9:108640415-111788150), separating out into three clusters across PC1: individuals homozygous for the common *mexicana* inversion allele (left), heterozygous individuals (middle) and individuals homozygous for the common maize inversion allele (right; includes all allopatric maize).

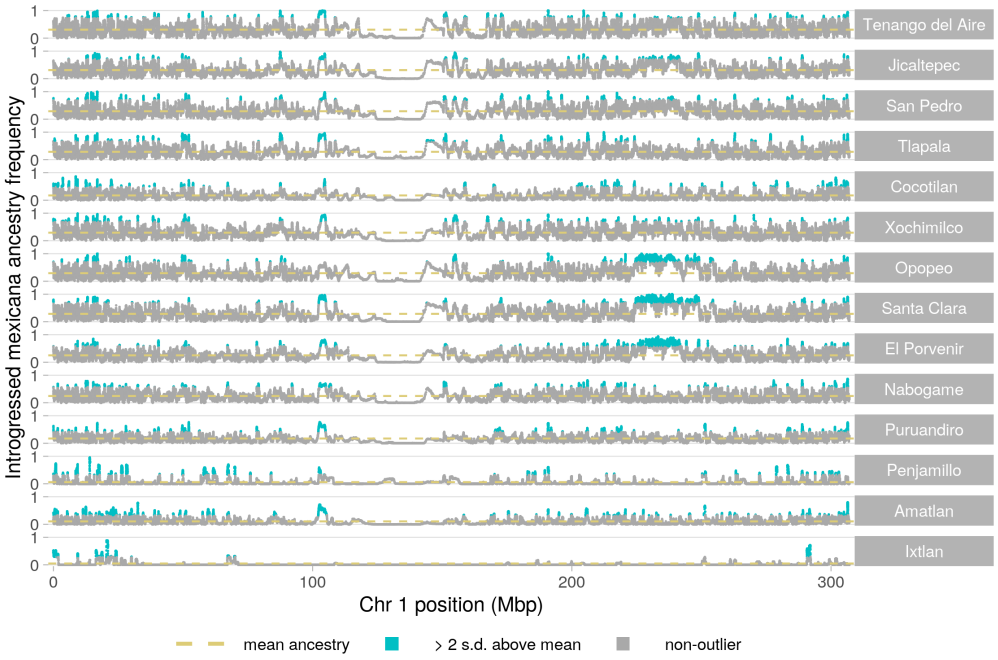

Fig S20. Introgression in maize landrace populations across chromosome 1

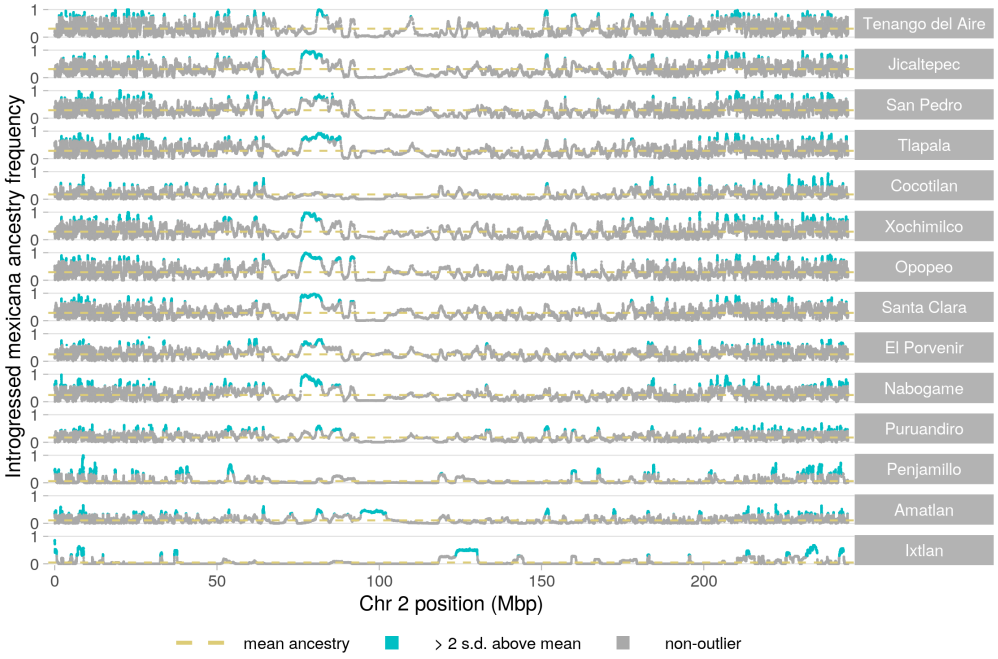

Fig S21. Introgression in maize landrace populations across chromosome 2

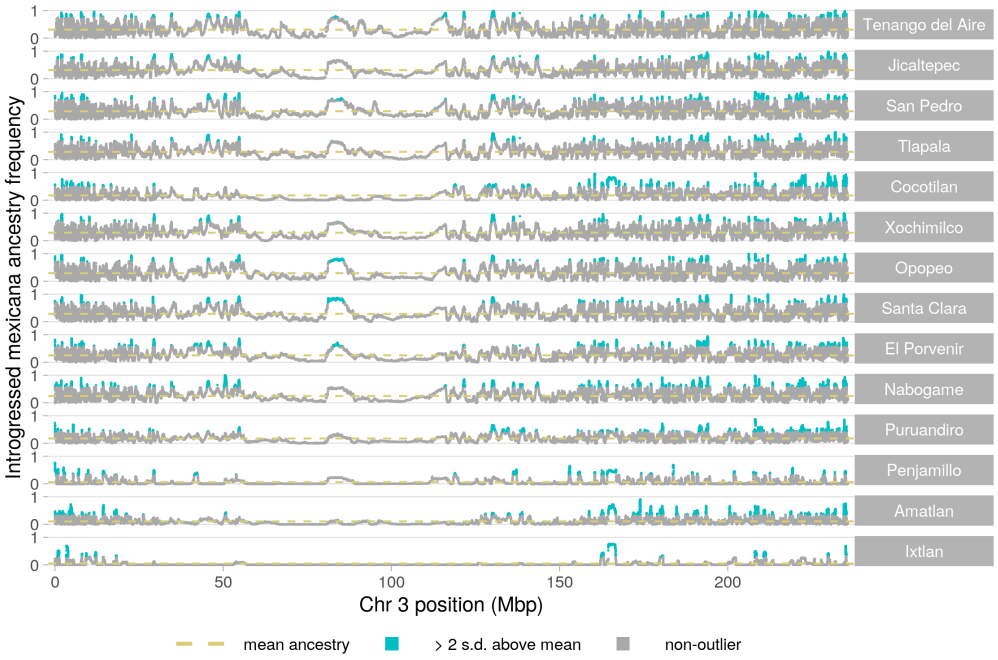

Fig S22. Introgression in maize landrace populations across chromosome 3

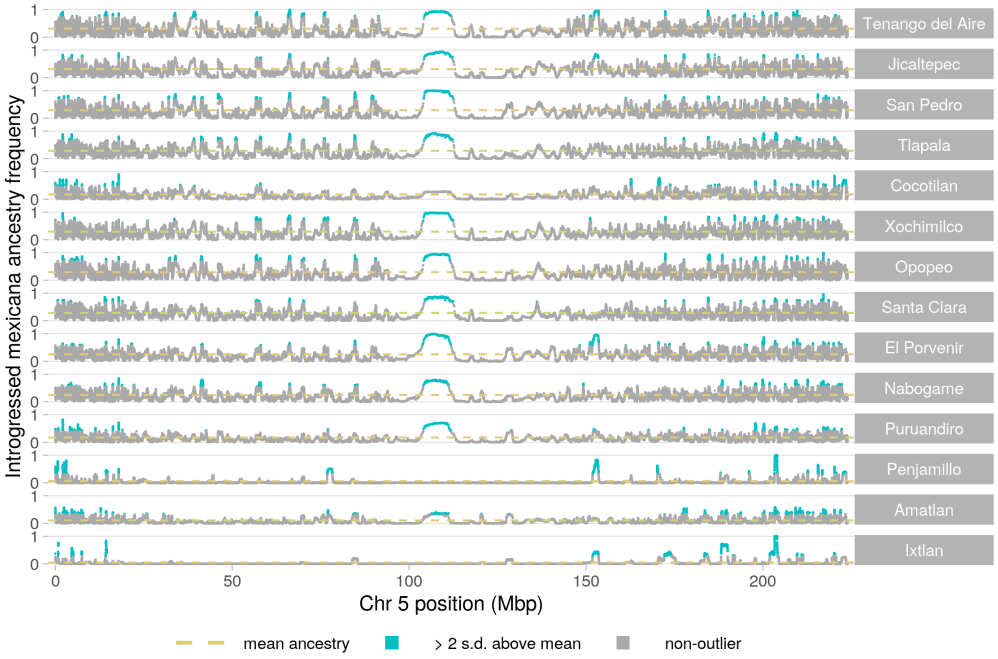

Fig S23. Introgression in maize landrace populations across chromosome 5

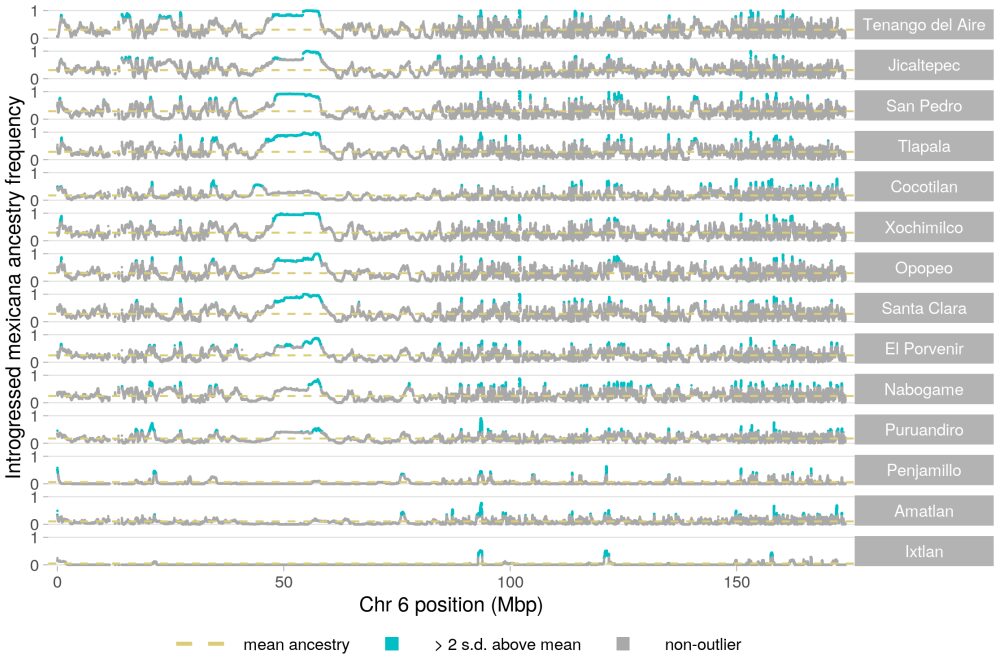

Fig S24. Introgression in maize landrace populations across chromosome 6

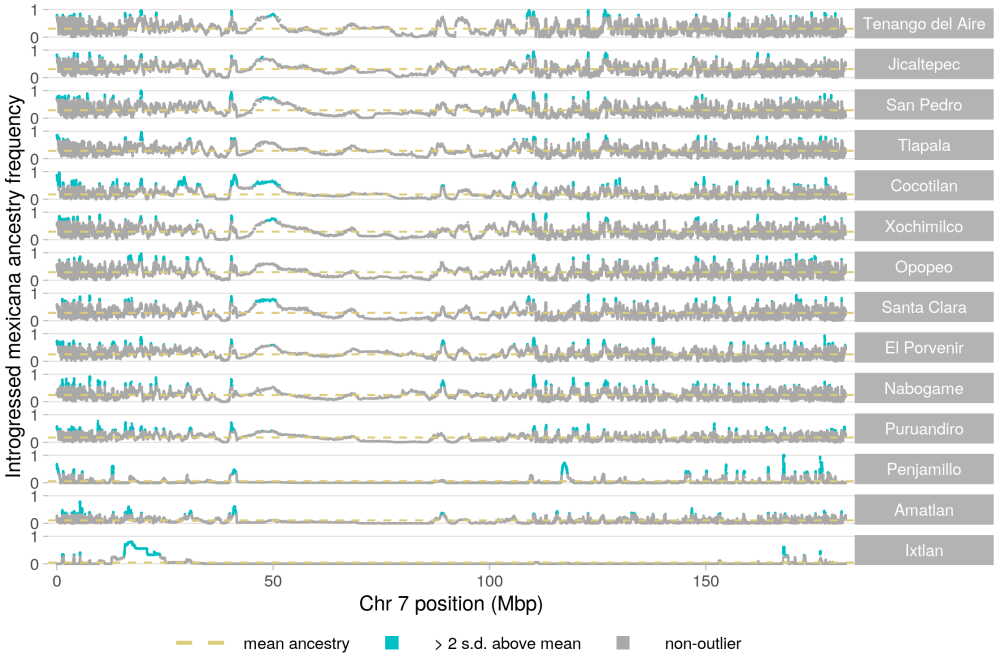

Fig S25. Introgression in maize landrace populations across chromosome 7

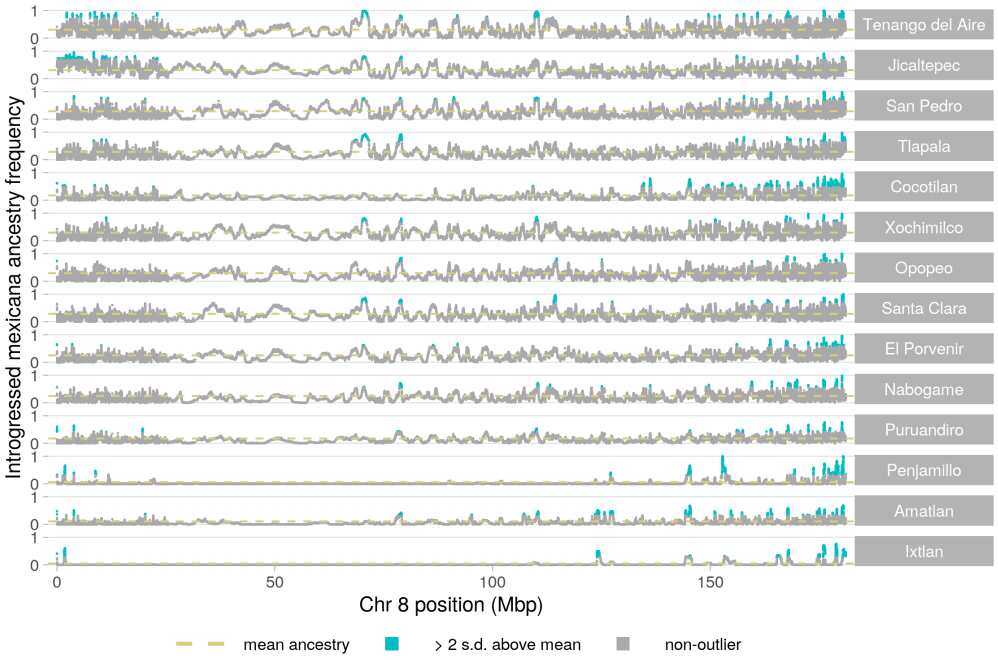

Fig S26. Introgression in maize landrace populations across chromosome 8

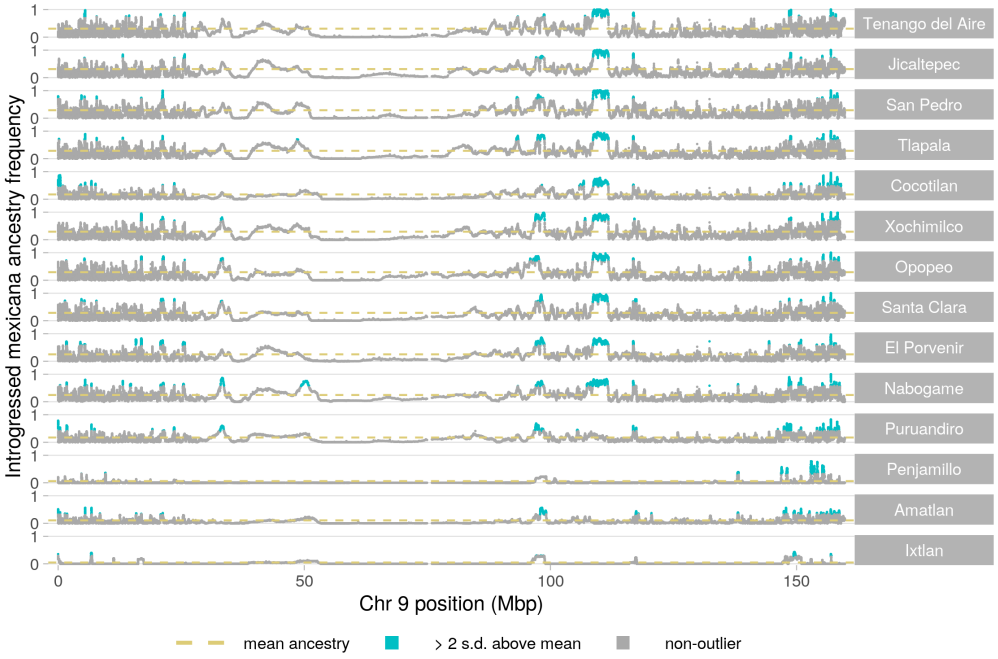

Fig S27. Introgression in maize landrace populations across chromosome 9

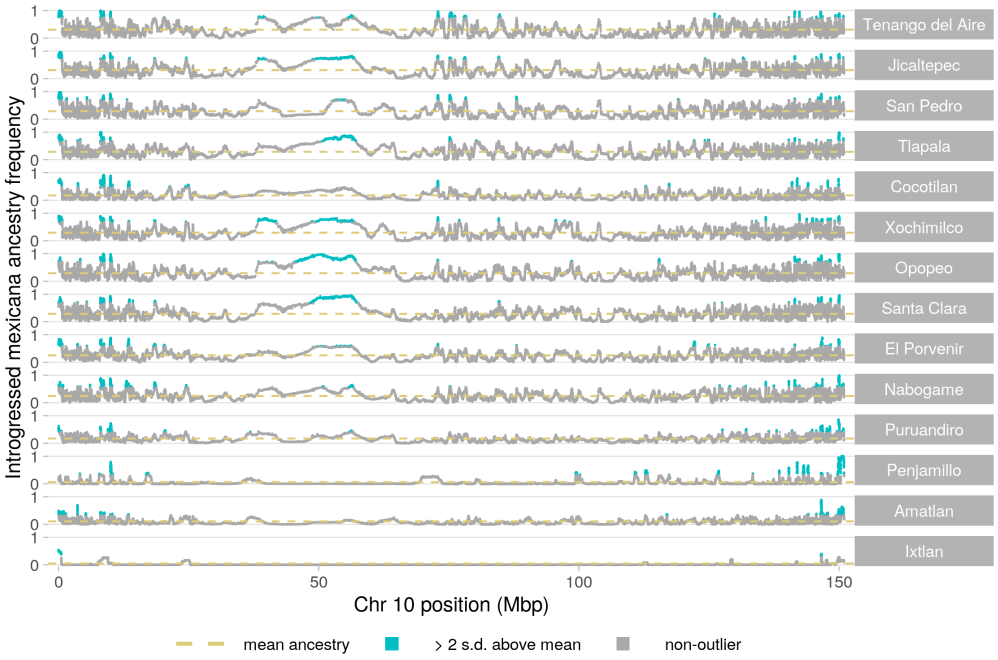

Fig S28. Introgression in maize landrace populations across chromosome 10

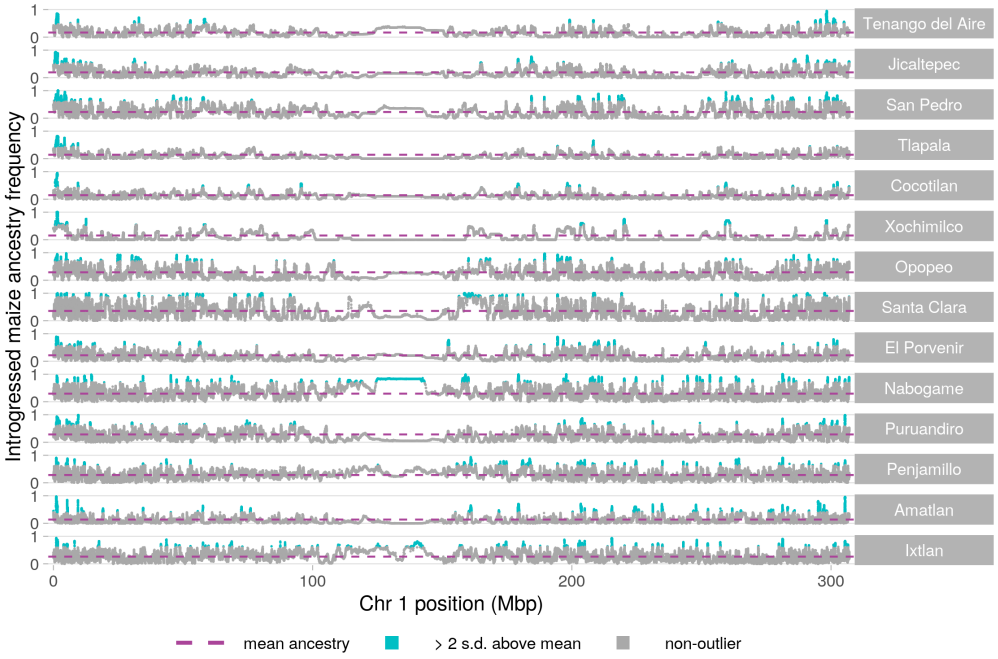

Fig S29. Introgression in *mexicana* populations across chromosome 1

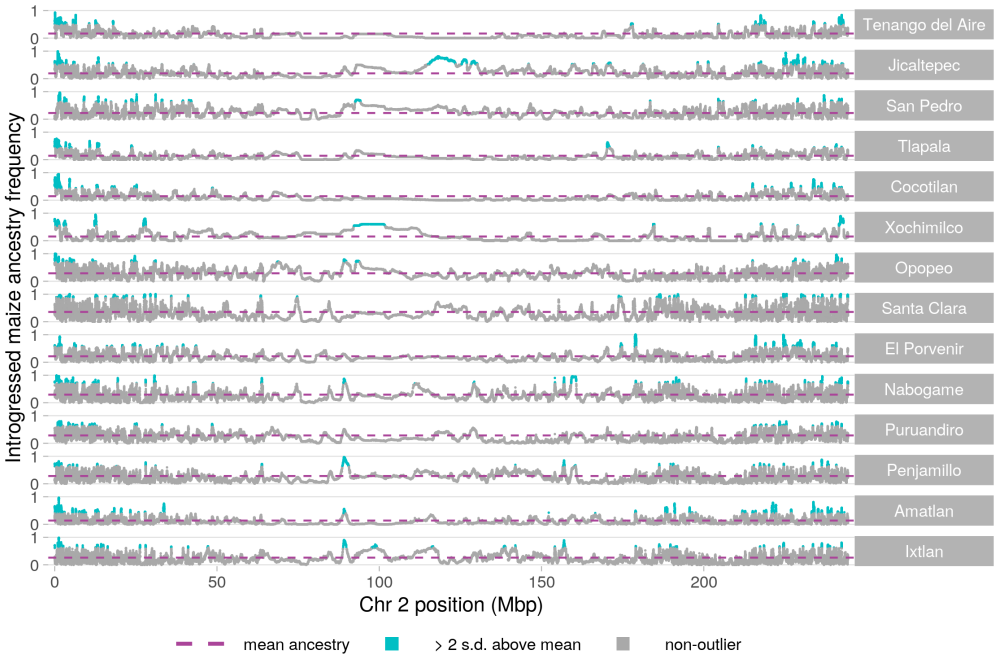

Fig S30. Introgression in *mexicana* populations across chromosome 2

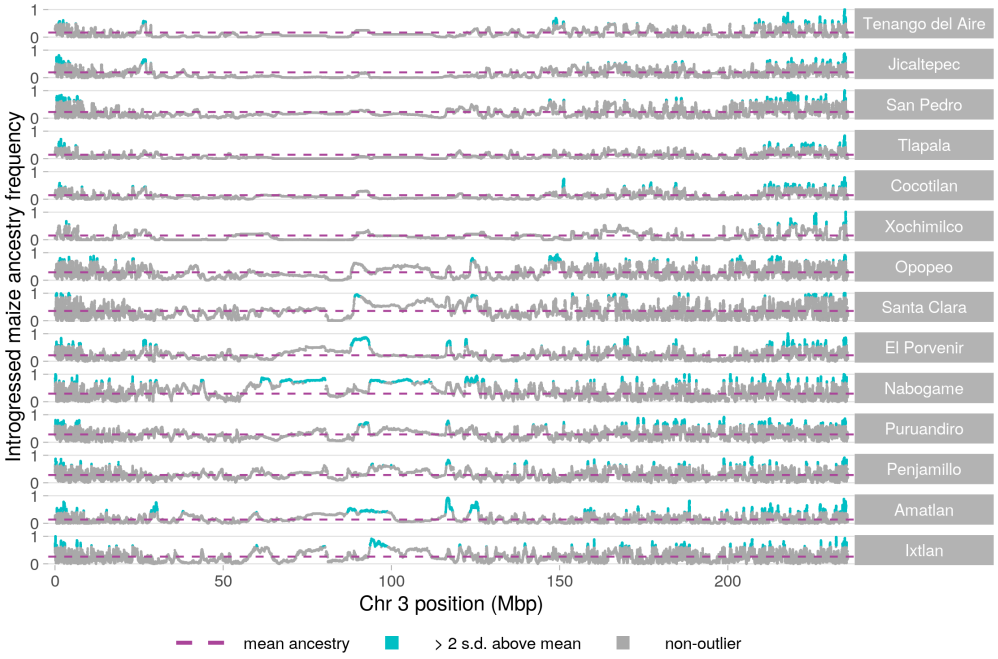

Fig S31. Introgression in *mexicana* populations across chromosome 3

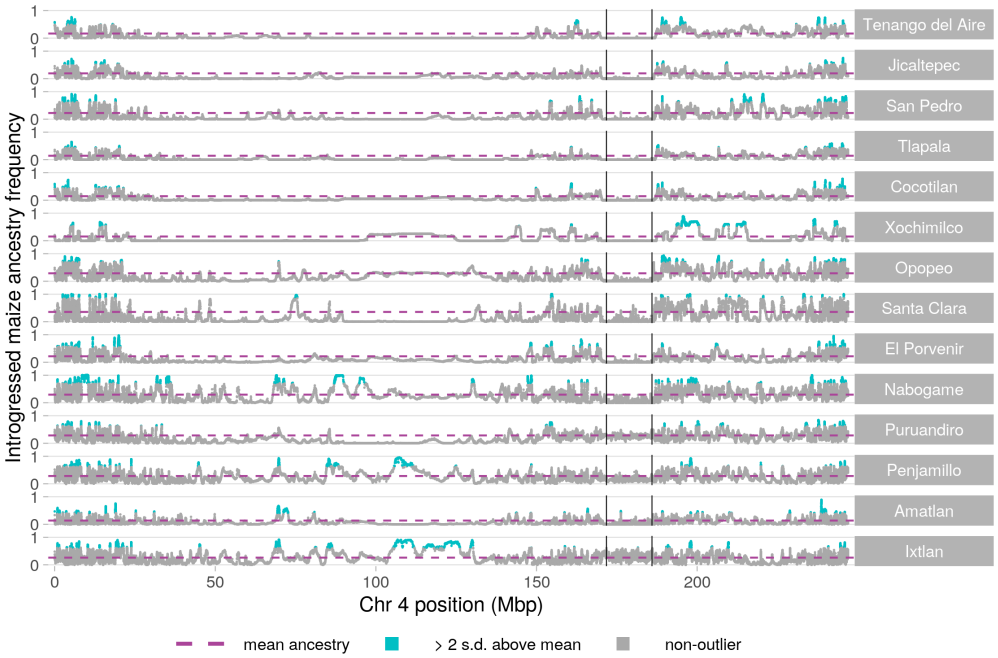

Fig S32. Introgression in *mexicana* populations across chromosome 4
Vertical lines indicate the coordinates for *Inv4m*.

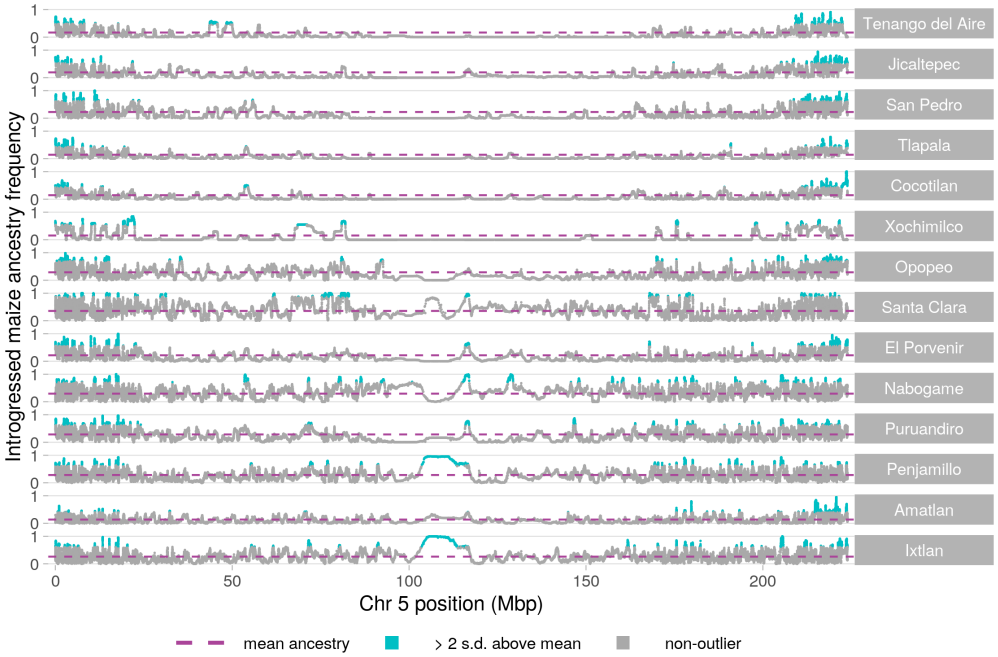

Fig S33. Introgression in *mexicana* populations across chromosome 5

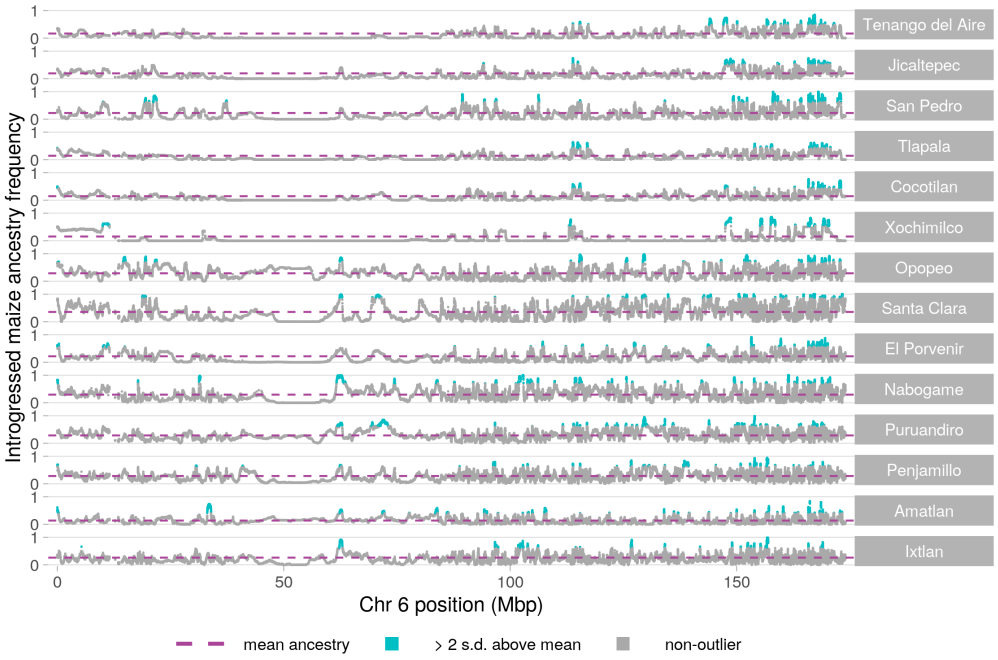

Fig S34. Introgression in *mexicana* populations across chromosome 6

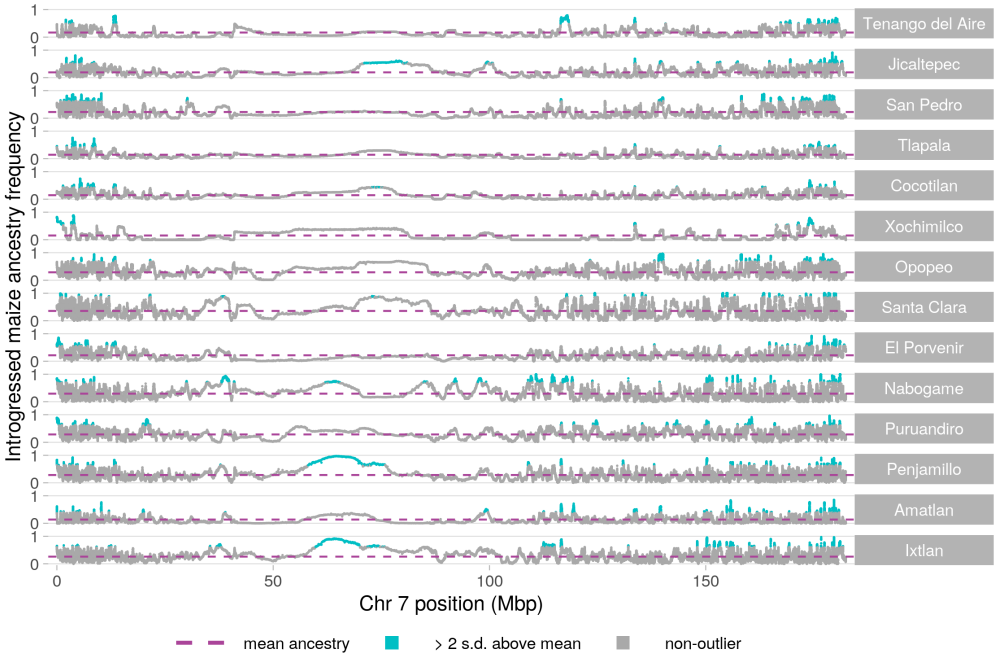

Fig S35. Introgression in *mexicana* populations across chromosome 7

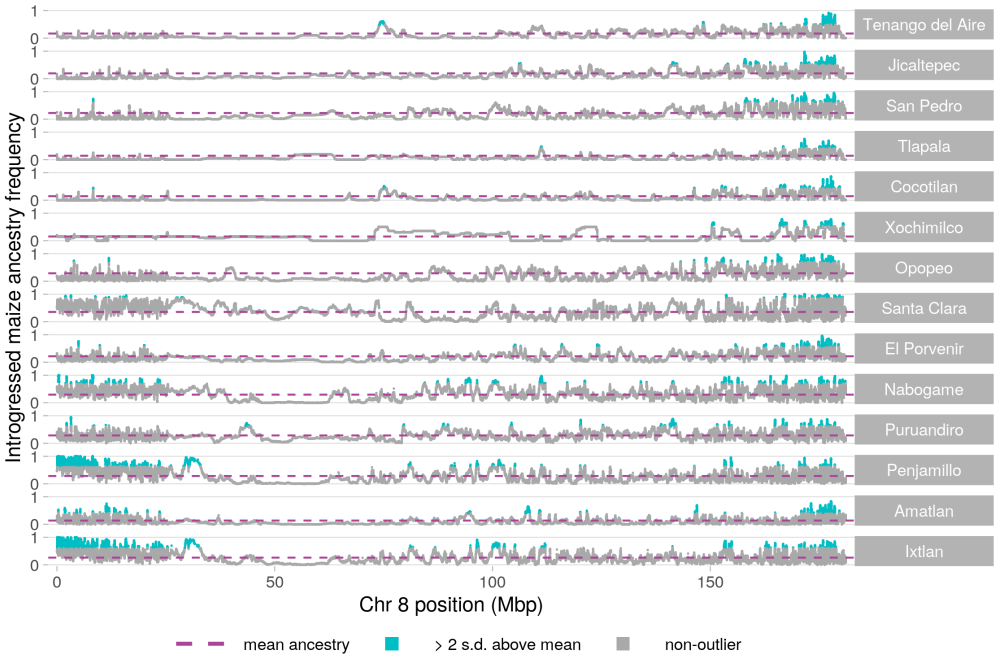

Fig S36. Introgression in *mexicana* populations across chromosome 8

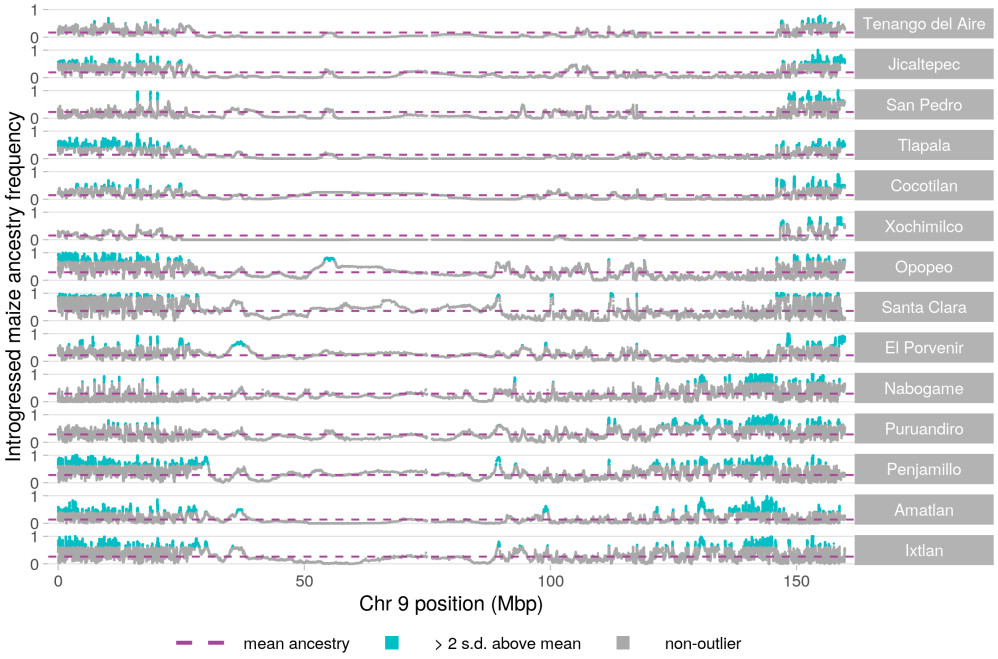

Fig S37. Introgression in *mexicana* populations across chromosome 9

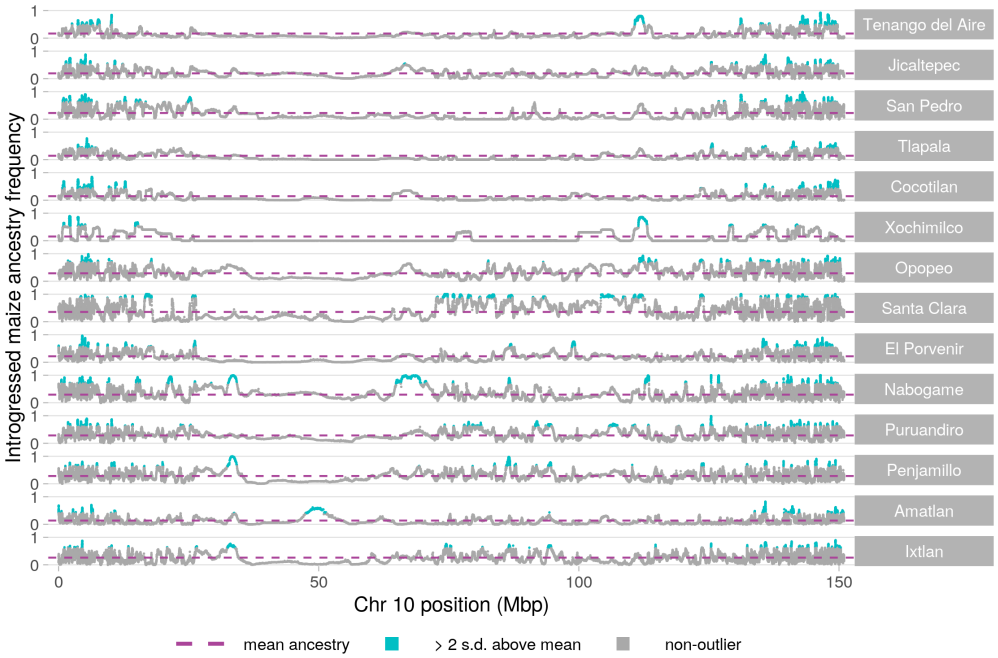

Fig S38. Introgression in *mexicana* populations across chromosome 10