bioRxiv preprint doi: https://doi.org/10.1101/2021.03.05.434040; this version posted September 21, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

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

Erin Calfee[1,2*], Daniel Gates[2,7], 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 Department of Biology, Geology, and Environmental Science, The University of Tennessee at Chattanooga, Tennessee, USA (# Current Address)
7 Checkerspot, Inc., Berkeley, California, USA (# Current Address)
¶ co-mentors.

*** erincalfee@gmail.com**

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, *mexicana* (*Zea mays* ssp. *mexicana*). 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 *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 into maize over 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 in maize 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

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.05.434040; this version posted September 21, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

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 14 Mb *Inv4m* on chromosome 4 and a novel 3 Mb inversion *Inv9f* surrounding the *macrohairless1* locus on chromosome 9. Most outlier loci with high *mexicana* introgression 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 populations.

Author Summary

When species expand their ranges, new encounters with diverse wild relatives can introduce deleterious genetic variation, but may also accelerate the colonization of novel environments by providing 'ready-made' genetic adaptations. Maize today is a global staple, far exceeding the original ecological niche of its wild progenitor. We show that gene flow from highland-endemic wild *mexicana* facilitated maize's range expansion from the valleys where it was domesticated to sites over 1500m in the mountains of Mexico. We find loci where *mexicana* ancestry has been repeatedly favored in highland maize populations. We also find loci (including a newly identified inversion) where *mexicana* ancestry increases steeply with elevation, providing evidence for adaptive trade-offs.

We additionally demonstrate selection against *mexicana* ancestry, especially near domestication genes. We sampled *mexicana* growing alongside maize fields, yet find little evidence that introgression is recent or locally-sourced genomewide or at adaptive loci. Rather, the majority of *mexicana* ancestry was introduced into maize over 1000 generations ago, and subsequently diverged and was sorted by selection in individual populations. These results add to our understanding of the effects of introgression on range expansions and adaptation.

Introduction

Interbreeding between partially diverged species or subspecies can result in admixed individuals with low fitness, e.g. due to hybrid incompatibilities [1–3]. Consistent with the view that hybridization is often deleterious, a growing number of species show evidence of pervasive selection against introgressed ancestry [4–13]. 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 [14]. Indeed, admixture has been linked to adaptive species radiations and/or rapid niche expansions in a number of natural systems, including mosquitoes [15], *Drosophila* [16], butterflies [9], cichlids [17], sunflowers [18], wild tomatoes [19] and yeast [20, 21]. In addition, introgression from wild relatives has facilitated the broad range expansions of multiple domesticated crops (reviewed in [22] and [23]), and gene flow from crops back into their wild relatives has in some cases opened up novel 'weedy' niches [24].

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 [25] and interbreeds with a number of wild relatives genetically adapted to distinct ecologies [26, 27]. 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

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.05.434040; this version posted September 21, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

domesticated in the Balsas River Valley in Mexico from a lowland-adapted sub-species of teosinte (*Zea mays* ssp. *parviglumis* [28]), which grows readily at sea level and lower elevations of the Sierra Madre del Sur [29]. In contrast, *Zea mays* ssp. *mexicana*, which diverged from *parviglumis* about 60 thousand years ago [30], 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 [31–35].

Maize was introduced as a crop to the mountains of Mexico around 6.2 thousand years ago [36], 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 [37, 38], including earlier flowering times for the shorter growing season [34], purple anthocyanin-based pigmentation which shields DNA from UV damage [39] and increases solar heat absorption [40], and macrohairs on the leaf and stem sheath, which are thought to increase herbivore defense [41] and/or heat maintenance in colder environments [42]. Earlier studies using 50K SNP-chip data for highland populations [43] or genomewide data for a small number of individuals [44, 45], have shown that highland maize populations 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. In addition, maize alleles introgressed into *mexicana* should be selected against because maize has accumulated genetic load from reduced population sizes during domestication [44] and because domestication traits generally reduce fitness in the wild [46–48], e.g. loss of disarticulation and effective seed dispersal [37].

In this study, we generate whole genome sequencing data to investigate genomic signatures of admixture and selection in paired sympatric maize and *mexicana* populations, sampled from 14 locations across an elevational gradient in Mexico. *Mexicana* was sampled from wild populations and maize was sampled from nearby fields where traditional cultivation methods and open pollination have resulted in populations with distinct local characteristics and high phenotypic and genetic diversity (often called maize 'landraces'). 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 [32, 37, 49–51]. 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 teosinte relatives, from sea-level up to 4000 meters [52]. Our sampled sites range from 1547-2600 meters in elevation, which

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.05.434040; this version posted September 21, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

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 individuals from 3 *mexicana* reference populations, totalling 348 low-coverage genomes (mean ∼1x). Two of these *mexicana* reference populations are documented to have no adjacent maize agriculture within the past 50 years, while a third higher elevation population (Amecameca) was chosen because it grows above the elevational range of *parviglumis*, and thus outside of the historical range of maize. For a maize reference population, we added 55 previously published high-coverage genomes from a population grown near Palmar Chico at 983 m [53, 54], well below the elevational range of *mexicana*. Because *parviglumis* is known to admix historically with both of our focal subspecies in Mexico, we also included 50 previously published high-coverage *parviglumis* genomes from the 'Mound' population at 1,008 m, also near Palmar Chico [53–55]. Completely allopatric reference populations are not available because maize has been grown at high density throughout Mexico across the full elevational range of both teosintes. *A priori*, gene flow from maize into *mexicana* is possible at Amecameca, and historically between maize and teosinte at all locations. We therefore assess for possible gene flow into each reference population below.

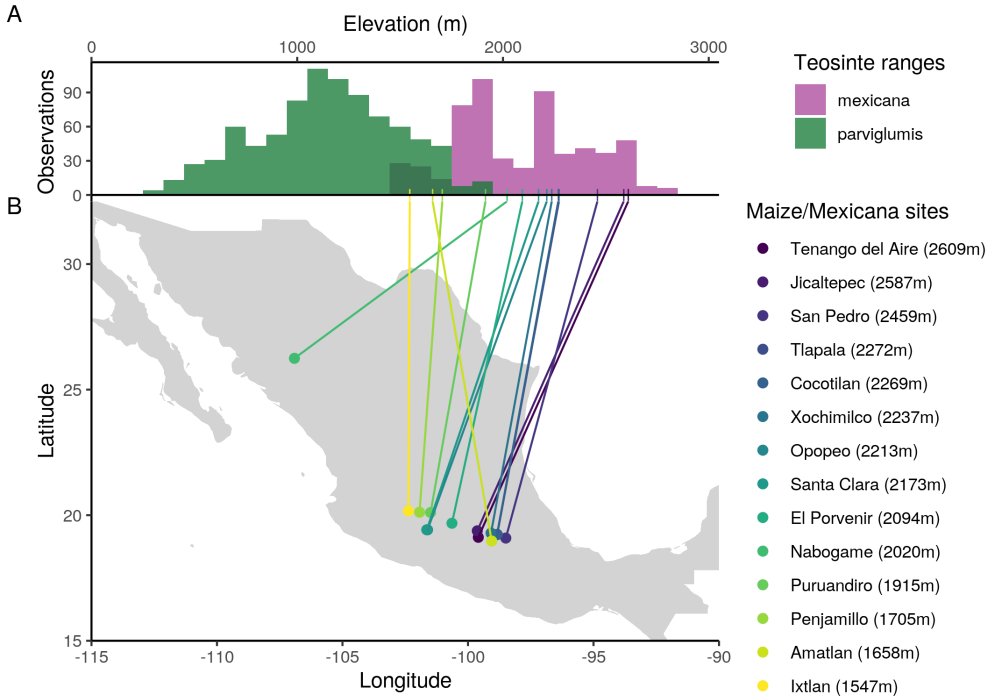

Fig 1. Sampled sympatric maize/*mexicana* populations compared to the distribution of teosintes (A) Elevational range of teosintes based on historical occurrence data (1842-2016) from [29]. *Parviglumis* and *mexicana* overlap at middle elevations (dark green) and maize today is grown across this entire elevational range. (B) Geographic location and elevation of contemporary sympatric maize and *mexicana* population pairs sampled across 14 sites in Mexico.

Principal components analysis of genetic diversity clearly separates maize and *mexicana*, with putative admixed individuals from sympatric populations having intermediate values along PC1. Additionally, PC2 provides evidence of gene flow from *parviglumis*, particularly into lower elevation *mexicana* populations (S1 Fig), which motivated us to analyse a 3-way admixture scenario in all subsequent analyses.

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.05.434040; this version posted September 21, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

To estimate genomewide ancestry proportions for each individual, we ran NGSAdmix [56] with K=3 genetic clusters and genotype likelihoods for all *mexicana*, maize, and *parviglumis* individuals. The three genetic clusters clearly map onto *mexicana*, maize and *parviglumis* ancestry, with no significant admixture in the lowland maize or *parviglumis* reference populations. We find minority *parviglumis* ancestry in the two lower-elevation reference *mexicana* populations, but no evidence of introgression from *parviglumis* or maize into the highest elevation population at Amecameca (Fig 2A).

Looking across samples from the 14 sympatric sites, we find a positive association between ancestry proportion and elevation (km), with higher *mexicana* ancestry at higher elevations in both sympatric maize ($\beta = 0.22, P = 1.01 \times 10^{-31}$) and sympatric *mexicana* ($\beta = 0.32, P = 2.02 \times 10^{-38}$) 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. While most populations have admixture proportions well-predicted by their elevation, outlier populations may be the result of recent colonization histories for some locations or adaptation to other environmental niches. Within teosintes, elevation is a major axis of niche separation between *parviglumis* (the ancestor of maize) and *mexicana* [50, 57], but genetic differentiation also correlates with soil nutrient content and at least four principal components constructed from climatic variables [33].

Origin and timing of introgression

If *mexicana* ancestry found in contemporary maize genomes facilitated maize's colonization of the highlands approximately 6.2 thousands years ago [36], 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 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 (HMM) based on read counts ([58]; see Materials and Methods). For each admixed population, this HMM simultaneously estimates local ancestry and, by optimizing the transition rate between different (hidden) ancestry states, the generations since admixture. We assumed a 3-way admixture scenario in which a founding *mexicana* population receives a pulse of *parviglumis* ancestry, then a pulse of maize ancestry.

Admixture between maize and *mexicana* is generally old, with median estimates of 1014 generations for sympatric maize populations and 509 generations for sympatric *mexicana* populations (S3 Fig). Parviglumis admixture timing estimates vary substantially across populations (median: 173, range: 29-1006). Because the HMM fits a single-pulse per ancestry 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) [59, 60] 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 *mexicana* ancestry 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

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.05.434040; this version posted September 21, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

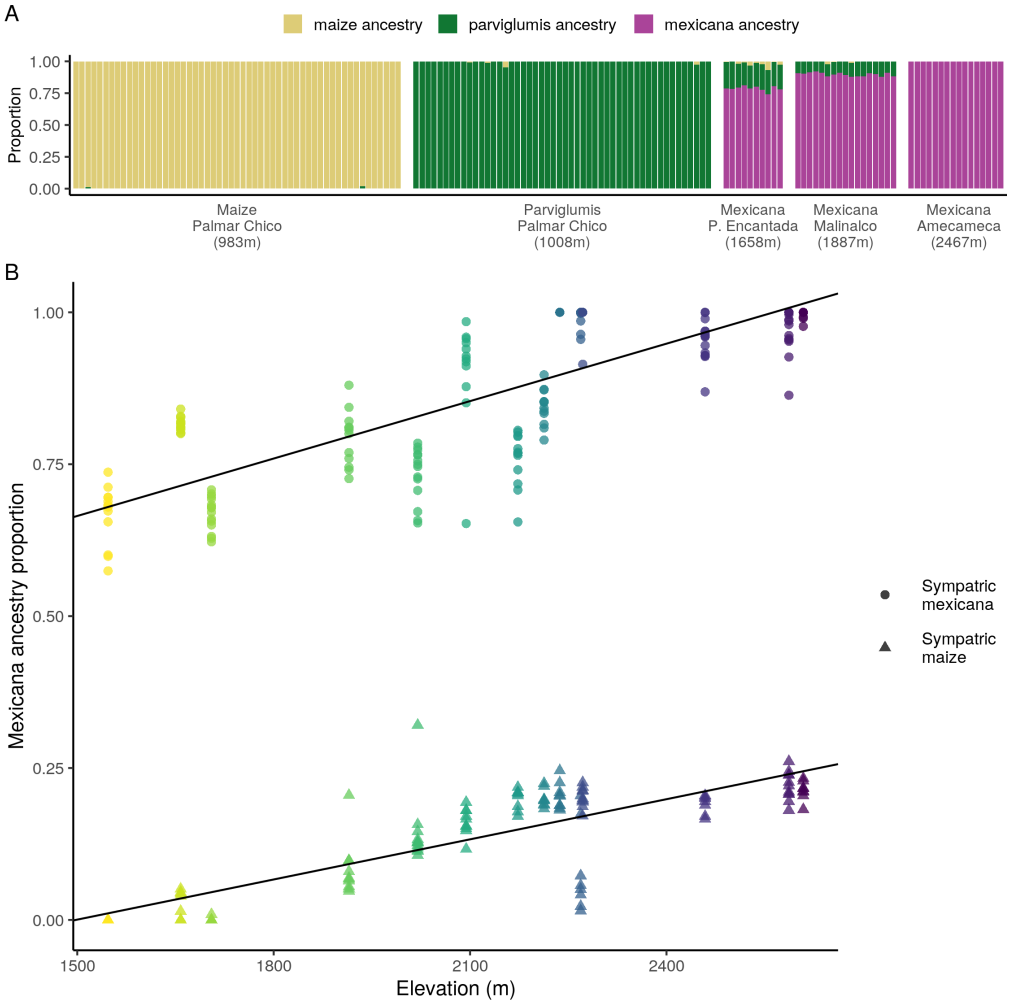

Fig 2. Distribution of *mexicana* ancestry by elevation (A) Genomewide ancestry estimates (NGSAdmix) for reference maize, *mexicana* and *parviglumis* individuals, grouped by sampling location. (B) Genomewide *mexicana* ancestry estimates (NGSAdmix) for sympatric maize and *mexicana* individuals (n =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

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* ancestry, once introgressed into a maize background, from its original source population(s) (S3 Fig). Additionally, *mexicana* ancestry tracts in maize have only slightly reduced genetic diversity (π, S4 Fig), 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 (Fig 3), more recent timing of admixture estimates (S3 Fig), and reduced genetic diversity (S4 Fig). These patterns

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.05.434040; this version posted September 21, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

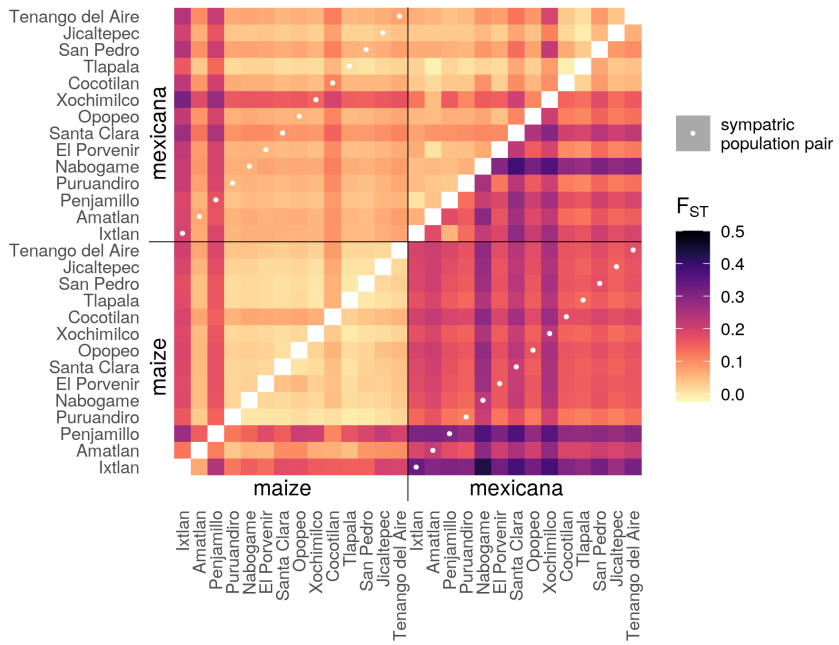

Fig 3. F_{ST} between ancestry tracts from different populations F_{ST} between each pair of populations for maize ancestry tracts are shown in the upper left triangle, while F_{ST} estimates for *mexicana* ancestry tracts are shown in the lower right triangle. Populations are sorted by subspecies, then elevation. Local sympatric maize-*mexicana* population pairs are highlighted with a white dot and do not show reduced F_{ST} relative to other (non-local) maize-*mexicana* comparisons. Additionally, introgressed *mexicana* ancestry shows low differentiation between maize populations (creating a light-colored maize block in the left corner of the lower right triangle) and no potential *mexicana* source populations show especially low F_{ST} with this block. Light coloring generally across the upper left triangle reflects the low differentiation within maize ancestry, providing little information to distinguish between potential maize ancestry sources.

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 [43]).

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 *mexicana* 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 adaptively introgressed haplotypes at a particular locus came 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 and S4).

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.05.434040; this version posted September 21, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

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 [4–13, 61]. 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=3 clusters, to estimate ancestry proportions for each quintile. We used a recombination map from maize [62], which is likely to be correlated with other *Zea* subspecies at least at the level of genomic quintiles. A limitation of this analysis, however, 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 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 ($\rho = 1.00,\ \mathrm{CI}_{95}[0.80,\ 1.00]$), reflecting the fact that introgression increases monotonically across quintiles. A similar analysis using f_4 statistics replicates this result (see Materials and Methods, S5 and S6 Figs). 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 positive interaction between recombination rate quintile and the slope of ancestry across elevation in sympatric maize (S4 Table). This is again consistent with low-recombination rate regions having a stronger effect of linked selection reducing *mexicana* ancestry, with higher elevation maize populations either experiencing larger amounts of gene flow or retaining more ancestry due to adaptive processes in high recombination regions.

Because recombination rate is positively correlated with gene density in *Zea* [63], 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 (S8 Fig, $\rho = -1.00,\ \mathrm{CI}_{95}[-1.00,\ -0.90]$).

In contrast, sympatric *mexicana* shows an unexpected negative relationship between recombination rate and introgression, with reduced maize ancestry in the highest recombination rate regions of the genome ($\rho = -1.00,\ \mathrm{CI}_{95}[-1.00,\ -0.90]$). Correlations with coding bp per cM and based on f_4 statistics corroborate this pattern (see S6 Fig). One explanation is that some portion of maize alleles are beneficial in a *mexicana* background.

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 [64] 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* introgression and recombination rates for 1 cM windows in sympatric maize (S9 Fig, $\rho = 0.011, \mathrm{CI}_{95}[-0.038,\ 0.061]$) and a positive rank correlation between maize introgression and recombination rate in sympatric *mexicana* ($\rho = 0.385,\ \mathrm{CI}_{95}[0.341,\ 0.428]$). Contrasting results between global and local ancestry methods could be a reflection of true evolutionary differences across different time

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.05.434040; this version posted September 21, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

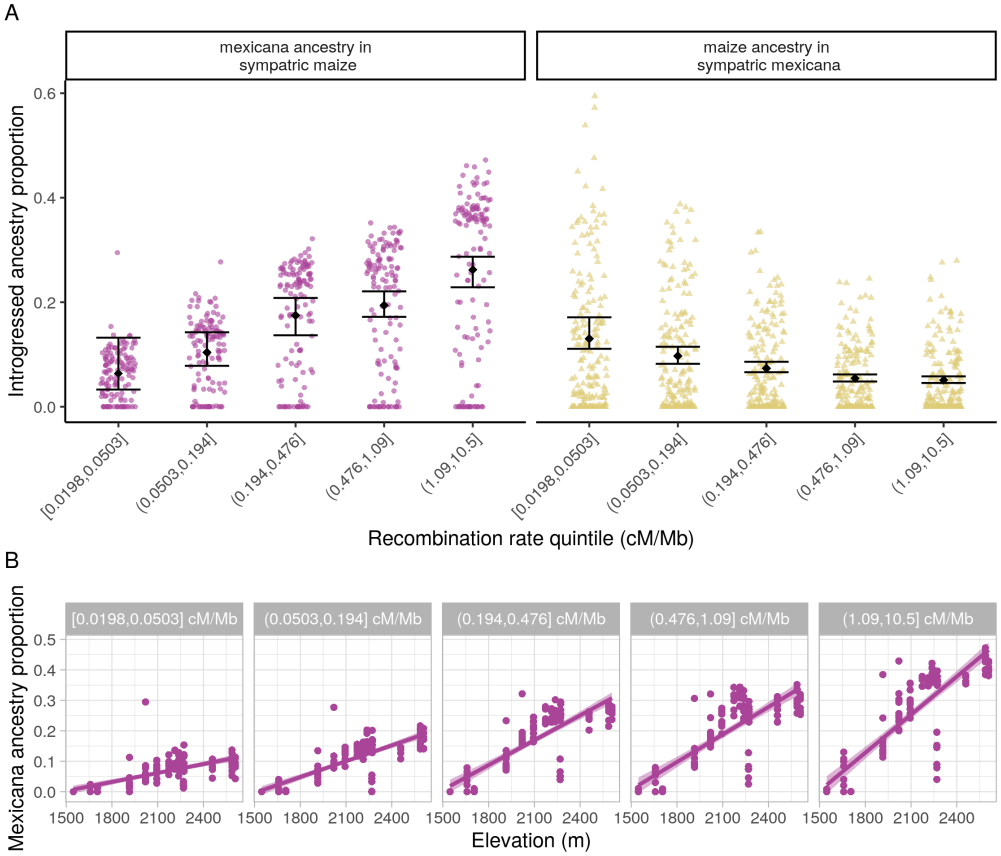

Fig 4. (A) Introgressed ancestry by recombination rate. Inferred average genomewide introgressed ancestry in sympatric maize and *mexicana* individuals (NGSAdmix K=3), by recombination rate quintiles. Group mean and 95% confidence interval based on bootstrap percentiles (n = 100) are depicted in black. Introgressed ancestry estimates for each individual are shown as points and points are jittered for better visualization. (B) Slope of *mexicana* ancestry introgressed into maize populations across elevation for each recombination rate quintile, based on NGSAdmix estimates. Each point is a sympatric maize individual and lines show the best-fit linear model for ancestry by elevation (with shaded 95% confidence interval), estimated separately for each quintile.

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 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*.

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.05.434040; this version posted September 21, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

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 (S4 Fig). This maintenance of diversity implies that selection at most peaks has favored multiple *mexicana* haplotypes, and hard sweeps for 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 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 (S10A Fig). To a lesser extent, we also observe sharing of high-introgression peaks for maize ancestry in sympatric *mexicana* populations (S10B Fig).

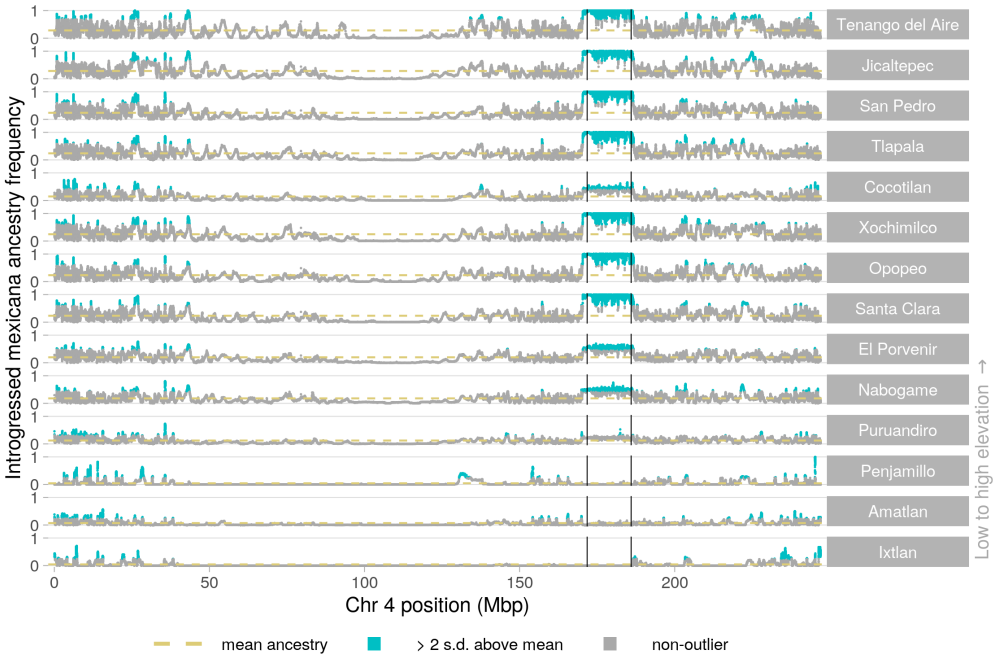

Fig 5. Introgression in maize populations across chromosome 4. Local introgressed *mexicana* ancestry frequency for each maize population compared to their genomewide mean. Populations are ordered from high to low elevation (top to bottom). High introgression peaks with more than 2 standard deviations 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 [50]). For local ancestry on other chromosomes and for maize introgression into sympatric *mexicana*, see Figs S11-S29.

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 Materials and Methods). These simulations preserve the ancestry variance across loci within populations and non-independence in ancestry between populations.

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.05.434040; this version posted September 21, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

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 their paired local *mexicana* population for these genomic regions. Patterns of F_{ST} between local sympatric pairs at local introgression peaks differed little from background F_{ST} (S30 Fig), 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 populations.

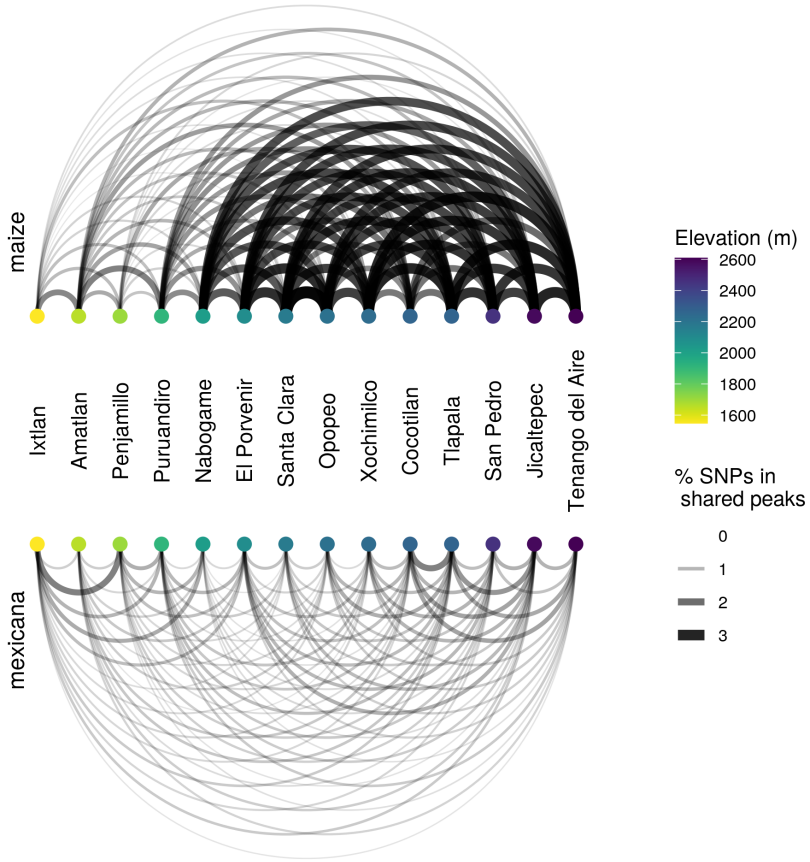

Fig 6. Introgression peaks shared across populations Networks for sympatric maize (top) and *mexicana* (bottom), where each node is a sampled population labelled by location and ordered by elevation. Edges connecting a pair of populations represent the percent of SNPs within shared ancestry peaks (introgressed 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. The relatively darker thicker lines connecting the high elevation maize populations (except for Cocotilan), indicate that these populations share high introgression peaks at especially high frequencies.

This lack of local adaptive introgression is perhaps surprising given the genetic

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.05.434040; this version posted September 21, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

structure in *mexicana* associated with different ecotypes [49] and evidence for local adaptation within teosinte across elevation [31, 50, 51]. 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 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 [37] 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* and maize introgression as well as steeper than expected slopes between *mexicana* ancestry and elevation (see S31 Fig for model fit).

We find several regions with high introgression in both directions that are unlikely to be explained by shared demographic history alone (Fig 7A). 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 additionally find evidence of adaptive *parviglumis* introgression into each subspecies (S32 Fig). 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 populations [43–45, 67]. This inversion maintains steep elevational clines within teosintes [50], overlaps QTLs for leaf pigmentation and macrohairs [42], and is associated with increased yield in maize at high elevations and decreased yield at low elevations [67], but has thus far eluded functional characterization of genes within the inversion [67].

Our second strongest association co-localizes with *macrohairless1* (*mhl1*), a locus on chromosome 9 that controls macrohair initiation on the leaf blade [65] and is associated with a major QTL explaining 52% of macrohair variation between high and low elevation teosinte mapping parents [42]. Within teosintes, populations of the lowland ancestor of maize, *parviglumis*, show convergent soft sweeps at the *mhl1* locus not shared by *mexicana* [32]. Macrohairs are characteristic highland phenotypes in teosinte and maize and are thought to confer adaptive benefits through insect defence and/or thermal insulation [41, 42]. 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 (S33 and S34 Figs). Additionally, we found evidence that the *mexicana*-type allele at the inversion segregates at low frequency within our lowland *parviglumis* reference population. Based on reduced diversity, the lowland maize/*parviglumis*-type allele at the inversion is likely derived (S6 Table). Thus *mexicana*-alleles at *Inv9f* could have been inherited by maize either through introgression or incomplete lineage sorting

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.05.434040; this version posted September 21, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

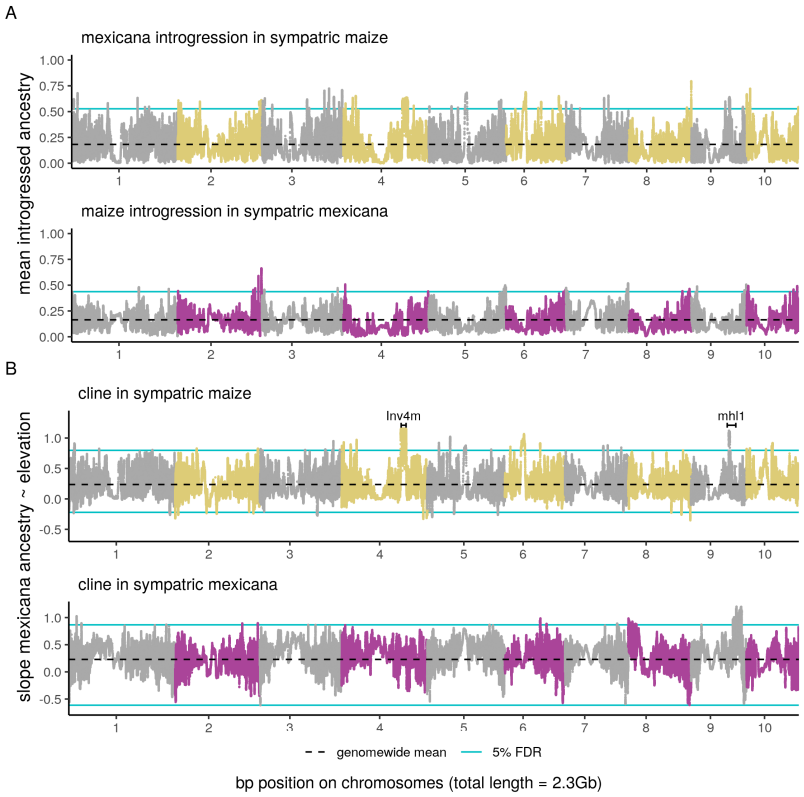

Fig 7. Genomewide scan for selection on introgressed ancestry. (A) Mean *mexicana* ancestry introgressed into sympatric maize populations and mean maize ancestry introgressed into sympatric *mexicana* populations. (B) Slope of *mexicana* ancestry proportion over a 1 km elevation gain in sympatric maize and *mexicana* populations. In both (A) and (B) the blue lines show the 5% false discovery rates, set using multi-variate normal simulations. Positions for *Inv4m* [50] and the *mhl1* locus [65] were converted to the maize reference genome v4 coordinates using Assembly Converter (ensembl.gramene.org). Chromosome numbers are placed at the centromere midpoint (approximate centromere positions are from [66]).

before selection pushed them to high frequency in highland 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* at the lower end of its elevational range.

Selection at candidate domestication genes

We hypothesized that domestication genes will be barriers to introgression bilaterally between maize and *mexicana* [43]. 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

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.05.434040; this version posted September 21, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

genome.

We examined introgression for a sample of 15 well-characterized domestication genes from the literature (see S7 Table), and compared them to the regions of the genome with the lowest 5% introgression of teosinte ancestry into sympatric maize and maize ancestry into sympatric *mexicana* ('introgression deserts'). A small but enriched subset of these domestication genes overlap with introgression deserts in sympatric maize (5, P < 0.001) and likewise in sympatric *mexicana* (5, P $= 0.001$).

Among these 15 domestication genes, we find that *teosinte branched1* (*tb1*), a key transciption factor that regulates branching vs. apical dominance [68, 69], overlaps introgression deserts in both maize and *mexicana*, consistent with *tb1*'s role at the top of the domestication regulatory hierarchy [70]. We also find evidence for reduced introgression into both maize and *mexicana* at *teosinte glume architecture1* (*tga1*) [71, 72], which is associated with 'naked' edible grains.

Another six domestication genes have low introgression in one direction only [73–77] (see S7 Table). Among these, *sugary1* (*su1*) in the starch pathway has low maize ancestry in *mexicana* but shows a steep increase in introgressed *mexicana* ancestry proportion with elevation in maize (+0.95 per km, $< 5\%$ FDR), which suggests this gene has pleiotropic effects on non-domestication traits in maize, with fitness trade-offs across elevation. *Sugary1* mutations modify the sweetness, nutrient content and texture of maize kernels (e.g. sweet corn), but also affect seed germination and emergence at cold temperatures [78], candidate pleiotropic effects that could be more deleterious at higher elevations.

The remaining seven domestication genes do not overlap introgression deserts in either subspecies despite evidence for their roles in domestication: *zfl2* (cob rank) [79–81], *pbf1* (storage protein synthesis) [82], *ba1* (plant architecture) [83], *ae1* (starch biosynthesis) [76], *ra1* and *ra2* (inflorescence architecture) [84, 85] and *ZmSh1-5.1+ZmSh1-5.2* (seed shattering) [75]. Despite evidence of introgression at many domestication loci, maize populations retain all of the classic domestication traits, and *mexicana* populations maintain 'wild' forms. Epistasis for domestication traits [47] 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. For example, the *mexicana* allele at *High PhosphatidylCholine 1* (*HPC1*) has recently been shown to reduce days to flowering and confer a fitness benefit in maize at higher elevations [86], and we find that the steepest clinal SNP within *HPC1* has a +0.79 *mexicana* ancestry proportion increase per km (FDR $< 10\%$), adding support for HPC1's role in adaptive earlier flowering at higher elevations in Mexican maize varieties (S35 Fig).

We tested more broadly for enriched selection within the flowering time pathway using a set of 849 candidate flowering time genes [87, 88] and a more stringent 5% FDR cutoff. No genes from the core flowering time pathway [87] and only 13/806 other candidate flowering time genes [88] (+/- 20kb) overlap outlier regions with steep increases in *mexicana* introgression with increasing elevation in sympatric maize, which matches expected overlap by chance (1.5%, P $= 0.93$). Thus steep clinal introgression patterns, indicative of strong fitness trade-offs across elevation, are the exception, not the rule, for flowering-time related genes. While maladaptive flowering times have

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.05.434040; this version posted September 21, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

strong fitness consequences, flowering time is also a highly polygenic trait [89], which may reduce the strength of selection at most individual genes to below what we can detect using steep ancestry clines. It is alternatively possible that *mexicana* alleles show adaptive benefits across the entire range sampled (moderate to high elevation), but we find that only 11/849 candidate flowering time genes overlap high mean *mexicana* introgression outliers at a 5% FDR (P = 0.95).

Conclusion

Intrinsic genetic incompatibilities and partial temporal isolation (offset flowering times) create an incomplete barrier to gene flow and F1 hybrids are commonly observed in the field [37], suggesting that hybridization is frequent and ongoing. Yet, we find little evidence of the effects of recent gene flow in either direction between sympatric maize-*mexicana* population pairs. One contributing factor may be selection on later generations: while hybrids are very challenging to identify and weed out from maize fields at early life stages, farmers can easily distinguish between maize and hybrids when choosing which cobs to plant for the next season. In the other direction, hybrids in most locations are expected to be partially temporally isolated from *mexicana* and hybrid seeds that do not disarticulate are farmer-dependent for successful dispersal and reproduction, although first-generation backcrosses to *mexicana* have been observed [37].

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 [47].

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. Indeed, timing of admixture estimates and differentiation of *mexicana* haplotypes within maize genomewide (F_{ST}) and at individual introgressed outlier loci (e.g. *Inv9f* at the *mhl1* locus (PCA)) suggest an ancient origin of introgressed haplotypes. 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.

Materials and Methods

Population sampling

We used maize and *mexicana* seed accessions sampled from locations across Mexico in 2008 [43] 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 [43]. In addition, we chose three population accessions to sequence as a *mexicana* reference panel: Puerta Encantada and

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.05.434040; this version posted September 21, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Malinalco were chosen because they have no record of contemporary maize agriculture nearby and a third population, Amecameca, was added as a complement to these two reference populations because it grows at a higher elevation, beyond the historical range of *parviglumis*.

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" [90]) 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 the three reference populations, with a total of 34.1x combined coverage. The remaining samples are maize and *mexicana* from paired sympatric populations, 262 of which have sufficient coverage for local ancestry inference ($\geq$ 0.5x, 6-12 per sympatric population, see S36 Fig). 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 [66] (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 [62] 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 (S37 Fig). From this map, we used *approx()* in R (v3.6.2 [91]) 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 [92]) and trimmed out adapter content from raw sequencing reads using the trimmomatic wrapper for snakemake (0.59.1/bio/trimmomatric/pe) [93]. We mapped trimmed reads to the maize reference genome using bwa mem (v0.7.17 [94]). We then sorted reads using SAMtools (v1.9 [95]), removed duplicates using picardtools (v2.7.1) MarkDuplicates and merged

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.05.434040; this version posted September 21, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

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 [96], which reduces the influence of bases with lower alignment quality.

Additional genomes from published sources

For a maize reference population, we used 55 previously published high-coverage genomes from a 'Tuxpeño' maize population grown at 983 m near Palmar Chico (NCBI: PRJNA616247 [53, 54]).

For a *parviglumis* reference population, we used 50 previously published high-coverage lowland individuals sampled from the 'Mound' population at 1,008 m near Palmar Chico [53–55] (NCBI: PRJNA616247, see S2 Table). These maize and parviglumis reference populations were sampled about 1 km apart from each other but maintain high F_{ST} and are isolated by differences in flowering time [54]. Both maize and *parviglumis* reference populations are allopatric to *mexicana*, growing well below its elevational range. We mapped and filtered reads for these reference maize and *parviglumis* 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 maize and *mexicana* genomes sequenced in this study and 105 high-coverage published genomes for reference maize and *parviglumis* described above. We used ANGSD (v0.932 [97]) 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' [96]) 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 [98]). In total, we identified 61,612,212 SNPs on the assembled chromosomes. In conjunction with SNP calling, we produced genotype likelihoods for each individual at these variant sites using the SAMtools GL method [95] implemented in ANGSD ('-GL 1 -doGlf 2').

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.05.434040; this version posted September 21, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

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 (∼4kb spacing) [99]. To first confirm that maize and *mexicana* ancestry form the major axis of genetic variation in our sample, we estimated the genetic covariance matrix between all individuals using PCAngsd (v0.98.2 [100]) 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 [56], using K = 3 clusters. Clusters clearly mapped onto the three reference groups, which we used to label the three ancestry components as 'maize', '*mexicana*' and '*parviglumis*'.

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 [58]). This method relies on allele counts from separate reference populations to estimate allele frequencies for each ancestry. Because some of our 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 in the top 10% tail for any of the three possible configurations of the population branch statistic between maize, *mexicana* and *parviglumis* reference populations, thereby enriching for SNPs that distinguish one reference population from the other two. We only considered SNPs with at least 44 reference maize, 12 reference *mexicana*, and 40 reference *parviglumis* individuals with sequencing coverage at a site. For these SNPs, we estimated reference population allele frequencies from angsd ('angsd -doMajorMinor 3 -GL 1 -baq 2 -doMaf 1'), then estimated pairwise F_{ST}'s using the Hudson 1992 estimator [101] as implemented by Bhatia et al. 2013 [102] to calculate the population branch statistic [103] (PBS equation from pg 8 of the supplement). We then calculated genetic distances between SNPs using the maize recombination map and filtered our enriched variants to have minimum 0.001 cM spacing between adjacent SNPs.

Running ancestry_hmm jointly infers local ancestry for each individual and the timing of admixture. This HMM method assumes a neutral demographic history in which a constant-size population receives a pulse of admixture t generations in the past, and finds the t that maximize the likelihood of the observed read counts and hidden local ancestry state across each admixed individual's genome. The timing of admixture defines the generations for possible meiotic recombination between ancestry tracts, and therefore scales the transition probabilities between hidden ancestry states. We ran ancestry_hmm under a three-way admixture model: a population is founded 10,000 generations ago by *mexicana*, then receives a pulse of *parviglumis* ancestry at t_{parv} (prior: 1000 generations ago, range: 0-10,000) and a pulse of maize ancestry at t_{maize} (prior: 100 generations ago; range: 0-10,000). Because these *Zea mays* subspecies are all annual grasses, generations can equivalently be interpreted as years since admixture. In addition to t_{parv} and t_{maize}, the HMM outputs the posterior probabilities for homozygous maize, homozygous *mexicana*, homozygous *parviglumis*, and heterozygous ancestry for each individual at every site. We analysed each sympatric maize and *mexicana* population separately, using the population's mean NGSAdmix global ancestry estimate as a prior for mixing proportions, an approximate effective population size (Ne)

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.05.434040; this version posted September 21, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

of 10,000 individuals, genetic positions for each SNP based on the maize linkage map, an estimated sequencing base error rate of 3×10^{-3}, and the three-way admixture model described above. 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 (t_{parv} and t_{maize}). 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 (S38 Fig).

To get a single point estimate for local ancestry at a site for an individual, we computed a sum of ancestry from the different possible ancestry states, weighted by their posterior probabilities, e.g. *mexicana* ancestry proportion = P(homozygous *mexicana*) + $^1/_2$P(heterozygous *mexicana*-maize) + $^1/_2$P(heterozygous *mexicana*-*parviglumis*). 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 local ancestry estimates from the HMM, we identified high-confidence homozygous ancestry tracts (posterior > 0.8). We filtered individual bams for reads that overlap these tracks and used the resulting filtered bams to calculate diversity within maize, *parviglumis*, 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 polarizing the unfolded site frequency spectrum in 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 [101] as parameterized in [102].

Effect of local recombination rate on introgressed ancestry

To estimate the effects of linked selection and recombination rate on genomewide introgression patterns, we compared introgressed ancestry estimates across genomic quintiles. Based on a 0.2 cM-resolution recombination map [62] 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') [104]. 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

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.05.434040; this version posted September 21, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

(once per quintile) with K=3 clusters. We assigned 'maize', '*mexicana*' and '*parviglumis*' labels to the ancestry clusters based on majority assignment to the respective 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. Genetic clusters could be unambiguously assigned to maize, *parviglumis*, and *mexicana* ancestries in all but 2 bootstrap replicates for the lowest recombination quintile and 3 for the highest coding density quintile, so we dropped this small number of replicates with unclear ancestry assignment from the bootstrap analysis. 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.

ii. 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. The f_4 statistic measures shared genetic drift (allelic covariance) between populations in a phylogeny, due to either shared branch lengths or admixture events in the evolutionary history relating these populations. Excess shared drift with one population from a pair of sister populations in the tree is a signature of admixture, analogous to the ABBA-BABA test [105], and a ratio of two f_4 statistics can be used to quantify the admixture proportion. Assuming the basic phylogenetic tree relating our reference populations (((*parviglumis*, maize), *mexicana*), *Tripsacum*) in S5 Fig, we can estimate α, the proportion of ancestry inherited from *mexicana* in a focal sympatric population, as follows [105, 106]:

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

The denominator of this statistic estimates the branch length leading to *parviglumis* and maize that separates these sister subspecies from *mexicana*; the full ratio estimates the proportion of this branch that separates sympatric population X from *parviglumis* and 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 reference *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 [107].

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

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.05.434040; this version posted September 21, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

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

iii. 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 maize, *mexicana* and *parviglumis* ancestry population frequencies marginally using a multivariate-normal null model, e.g.

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

where $\vec{\alpha}$ is the vector of mean *mexicana* ancestry frequencies genomewide for each sympatric population and K is the marginal *mexicana* ancestry variance-covariance matrix relating these 14 populations and estimated from the empirical data. 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 each focal ancestry marginally 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

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.05.434040; this version posted September 21, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

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 introgressed 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 introgressed ancestry using the Benjamini-Hochberg method [108] 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% genomewide of teosinte introgression into sympatric maize (or, separately, maize introgression into sympatric *mexicana*). We then looked up v4 coordinates on Ensembl.org for genes associated with maize domestication in the literature (S7 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 [87], and a broader list of 905 flowering time candidate genes [87, 88]. 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 [109]. 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.

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.05.434040; this version posted September 21, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Analysis pipeline and data visualization

We constructed and ran bioinformatics pipelines using snakemake (v.5.17.0 [110]) within a python conda environment (v3.6). We analyzed and visualized data in R (v3.6.2 [91]) using the following major packages: tidyverse (v1.3.0 [111]), viridis (v0.5.1 [112]), bedr (v1.0.7 [104]), boot (v.1.3_25 [113, 114]), gridExtra (v2.3 [115]), ggupset (v0.3.0 [116]) and tidygraph (1.2.0 [117]). 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).

Acknowledgments

We thank Pesach Lubinsky for collecting the seeds sequenced in this study. We also want to thank Michael Turelli, the Coop and Ross-Ibarra labs, and the HILO and Zeavolution working groups for helpful feedback on this work.

Funding

This work was funded by the the Division of Integrative Organismal Systems from the National Science Foundation, www.nsf.gov (NSF No. 1546719, awarded to JRI and GC), the National Institute of General Medical Sciences of the National Institutes of Health, www.nigms.nih.gov (NIH R01 GM108779 and R35 GM136290, awarded to GC), and the University of California, Davis, https://ucdavis.edu (Loomis Graduate Fellowship in Agronomy, awarded to EC). The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Data Availability

Raw Illumina sequence data generated by this study is available through the NCBI Short Read Archive, PRJNA657016 (maize/*mexicana*) and SRR7758238 (*Tripsacum*). All relevant metadata is found within the Supporting Information files. Access to previously published genomic resources used in this study: B73 maize reference genome v4 (Gramene), recombination map from Ogut et al. 2015 (https://www.panzea.org/publications), maize reference panel sequences (NCBI SRA PRJNA616247), and *parviglumis* reference panel sequences (NCBI SRA PRJNA616247). Local ancestry posterior probability files, ancestry_hmm input files, genomewide ancestry estimates, and summarised population allele frequency data are accessible via Figshare (https://doi.org/10.6084/m9.figshare.16641799). Scripts are available at https://github.com/ecalfee/hilo.

References

1. Powell DL, García-Olazábal M, Keegan M, Reilly P, Du K, Díaz-Loyo AP, et al. Natural hybridization reveals incompatible alleles that cause melanoma in swordtail fish. Science. 2020;368(6492):731–736. doi:10.1126/science.aba5216.

2. Zuellig MP, Sweigart AL. Gene duplicates cause hybrid lethality between sympatric species of Mimulus. PLOS Genetics. 2018;14(4):e1007130. doi:10.1371/journal.pgen.1007130.

3. Presgraves DC. The molecular evolutionary basis of species formation. Nature Reviews Genetics. 2010;11(33):175–180. doi:10.1038/nrg2718.

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.05.434040; this version posted September 21, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

4. Brandvain Y, Kenney AM, Flagel L, Coop G, Sweigart AL. Speciation and introgression between Mimulus nasutus and Mimulus guttatus. PLoS Genetics. 2014;10(6):e1004410.

5. Aeschbacher S, Selby JP, Willis JH, Coop G. Population-genomic inference of the strength and timing of selection against gene flow. Proceedings of the National Academy of Sciences. 2017;114(27):7061–7066. doi:10.1073/pnas.1616755114.

6. Kenney AM, Sweigart AL. Reproductive isolation and introgression between sympatric Mimulus species. Molecular Ecology. 2016;25(11):2499–2517.

7. Nelson TC, Stathos AM, Vanderpool DD, Finseth FR, Yuan Yw, Fishman L. Ancient and recent introgression shape the evolutionary history of pollinator adaptation and speciation in a model monkeyflower radiation (Mimulus section Erythranthe). PLOS Genetics. 2021;17(2):e1009095. doi:10.1371/journal.pgen.1009095.

8. Martin SH, Davey JW, Salazar C, Jiggins CD. Recombination rate variation shapes barriers to introgression across butterfly genomes. PLOS Biology. 2019;17(2):e2006288. doi:10.1371/journal.pbio.2006288.

9. Edelman NB, Frandsen PB, Miyagi M, Clavijo B, Davey J, Dikow RB, et al. Genomic architecture and introgression shape a butterfly radiation. Science. 2019;366(6465):594–599. doi:10.1126/science.aaw2090.

10. Sankararaman S, Mallick S, Dannemann M, Prüfer K, Kelso J, Pääbo S, et al. The genomic landscape of Neanderthal ancestry in present-day humans. Nature. 2014;507(74927492):354–357. doi:10.1038/nature12961.

11. Harris K, Nielsen R. The Genetic Cost of Neanderthal Introgression. Genetics. 2016;203(2):881–891.

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

13. Schumer M, Xu C, Powell DL, Durvasula A, Skov L, Holland C, et al. Natural selection interacts with recombination to shape the evolution of hybrid genomes. Science. 2018;360(6389):656–660.

14. Pfennig KS, Kelly AL, Pierce AA. Hybridization as a facilitator of species range expansion. Proceedings of the Royal Society B: Biological Sciences. 2016;283(1839):20161329. doi:10.1098/rspb.2016.1329.

15. Fontaine MC, Pease JB, Steele A, Waterhouse RM, Neafsey DE, Sharakhov IV, et al. Mosquito genomics. Extensive introgression in a malaria vector species complex revealed by phylogenomics. Science (New York, NY). 2015;347(6217):1258524. doi:10.1126/science.1258524.

16. Suvorov A, Kim BY, Wang J, Armstrong EE, Peede D, D'Agostino ERR, et al. Widespread introgression across a phylogeny of 155 Drosophila genomes. bioRxiv. 2021; p. 33. doi:https://doi.org/10.1101/2020.12.14.422758.

17. Martin CH, Cutler JS, Friel JP, Dening Touokong C, Coop G, Wainwright PC. Complex histories of repeated gene flow in Cameroon crater lake cichlids cast doubt on one of the clearest examples of sympatric speciation. Evolution. 2015;69(6):1406–1422.

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.05.434040; this version posted September 21, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

18. Rieseberg LH, Raymond O, Rosenthal DM, Lai Z, Livingstone K, Nakazato T, et al. Major Ecological Transitions in Wild Sunflowers Facilitated by Hybridization. Science. 2003;301(5637):1211–1216. doi:10.1126/science.1086949.

19. Pease JB, Haak DC, Hahn MW, Moyle LC. Phylogenomics Reveals Three Sources of Adaptive Variation during a Rapid Radiation. PLOS Biology. 2016;14(2):e1002379. doi:10.1371/journal.pbio.1002379.

20. Eberlein C, Hénault M, Fijarczyk A, Charron G, Bouvier M, Kohn L, et al. Hybridization is a recurrent evolutionary stimulus in wild yeast speciation. Nature Communications. 2019;10(1):923–923. doi:10.1038/s41467-019-08809-7.

21. Tusso S, Nieuwenhuis BPS, Sedlazeck FJ, Davey JW, Jeffares DC, Wolf JBW. Ancestral Admixture Is the Main Determinant of Global Biodiversity in Fission Yeast. Molecular Biology and Evolution. 2019;36(9):1975–1989. doi:10.1093/molbev/msz126.

22. Purugganan MD. Evolutionary Insights into the Nature of Plant Domestication. Current Biology. 2019;29(14):R705–R714. doi:10.1016/j.cub.2019.05.053.

23. Janzen GM, Wang L, Hufford MB. The extent of adaptive wild introgression in crops. New Phytologist. 2019;221(3):1279–1288. doi:https://doi.org/10.1111/nph.15457.

24. Ellstrand NC, Meirmans P, Rong J, Bartsch D, Ghosh A, de Jong TJ, et al. Introgression of Crop Alleles into Wild or Weedy Populations. Annual Review of Ecology, Evolution, and Systematics. 2013;44(1):325–345. doi:10.1146/annurev-ecolsys-110512-135840.

25. Hake S, Ross-Ibarra J. Genetic, evolutionary and plant breeding insights from the domestication of maize. eLife. 2015;4:e05861. doi:10.7554/eLife.05861.

26. Hufford MB, Bilinski P, Pyhäjärvi T, Ross-Ibarra J. Teosinte as a model system for population and ecological genomics. Trends in Genetics. 2012;28(12):606–615. doi:10.1016/j.tig.2012.08.004.

27. Mammadov J, Buyyarapu R, Guttikonda SK, Parliament K, Abdurakhmonov IY, Kumpatla SP. Wild Relatives of Maize, Rice, Cotton, and Soybean: Treasure Troves for Tolerance to Biotic and Abiotic Stresses. Frontiers in Plant Science. 2018;9. doi:10.3389/fpls.2018.00886.

28. Piperno DR, Ranere AJ, Holst I, Iriarte J, Dickau R. 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. 2009;106(13):5019–5024. doi:10.1073/pnas.0812525106.

29. González JdJS, Corral JAR, García GM, Ojeda GR, Larios LDlC, Holland JB, et al. Ecogeography of teosinte. PLOS ONE. 2018;13(2):e0192676. doi:10.1371/journal.pone.0192676.

30. Ross-Ibarra J, Tenaillon M, Gaut BS. Historical Divergence and Gene Flow in the Genus Zea. Genetics. 2009;181(4):1399–1413. doi:10.1534/genetics.108.097238.

31. Fustier MA, Martínez-Ainsworth NE, Aguirre-Liguori JA, Venon A, Corti H, Rousselet A, et al. Common gardens in teosintes reveal the establishment of a syndrome of adaptation to altitude. PLOS Genetics. 2019;15(12):e1008512. doi:10.1371/journal.pgen.1008512.

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.05.434040; this version posted September 21, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

32. Fustier MA, Brandenburg JT, Boitard S, Lapeyronnie J, Eguiarte LE, Vigouroux Y, et al. Signatures of local adaptation in lowland and highland teosintes from whole-genome sequencing of pooled samples. Molecular Ecology. 2017;26(10):2738–2756. doi:https://doi.org/10.1111/mec.14082.

33. Aguirre-Liguori JA, Gaut BS, Jaramillo-Correa JP, Tenaillon MI, Montes-Hernández S, García-Oliva F, et al. 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. 2019;28(11):2814–2830. doi:https://doi.org/10.1111/mec.15098.

34. Rodríguez JG, Sánchez G, Baltazar BM, Cruz LLDl, Santacruz-Ruvalcaba F, Ron PJ, et al. Characterization of floral morphology and synchrony among Zea species in Mexico. Maydica. 2006; p. 383–398.

35. Piazena H. The effect of altitude upon the solar UV-B and UV-A irradiance in the tropical Chilean Andes. Solar Energy. 1996;57(2):133–140. doi:10.1016/S0038-092X(96)00049-7.

36. Piperno DR, Flannery KV. 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. 2001;98(4):2101–2103. doi:10.1073/pnas.98.4.2101.

37. Wilkes HG. Teosinte: The Closest Relative of Maize. The Bussey Institute of Harvard University; 1967.

38. Wilkes HG. Hybridization of maize and teosinte, in Mexico and Guatemala and improvement of maize. Economic Botany. 1977;31(3):254–293.

39. Stapleton AE, Walbot V. Flavonoids Can Protect Maize DNA from the Induction of Ultraviolet Radiation Damage. Plant Physiology. 1994;105(3):881–889. doi:10.1104/pp.105.3.881.

40. Barthakur N. Temperature differences between two pigmented types of corn plants. International Journal of Biometeorology. 1974;18(1):70–75. doi:10.1007/BF01450666.

41. Moya-Raygoza G. 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. 2016;109(5):737–743. doi:10.1093/aesa/saw049.

42. Lauter N, Gustus C, Westerbergh A, Doebley J. The Inheritance and Evolution of Leaf Pigmentation and Pubescence in Teosinte. Genetics. 2004;167(4):1949–1959. doi:10.1534/genetics.104.026997.

43. Hufford MB, Lubinksy P, Pyhäjärvi T, Devengenzo MT, Ellstrand NC, Ross-Ibarra J. The genomic signature of crop-wild introgression in maize. PLoS Genetics. 2013;9(5):e1003477.

44. Wang L, Beissinger TM, Lorant A, Ross-Ibarra C, Ross-Ibarra J, Hufford MB. The interplay of demography and selection during maize domestication and expansion. Genome biology. 2017;18(1):215.

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.05.434040; this version posted September 21, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

45. Gonzalez-Segovia E, Pérez-Limon S, Cíntora-Martínez GC, Guerrero-Zavala A, Janzen GM, Hufford MB, et al. Characterization of introgression from the teosinte Zea mays ssp. mexicana to Mexican highland maize. PeerJ. 2019;7:e6815. doi:10.7717/peerj.6815.

46. Meyer RS, Purugganan MD. Evolution of crop species: genetics of domestication and diversification. Nature Reviews Genetics. 2013;14(1212):840–852. doi:10.1038/nrg3605.

47. Stitzer MC, Ross-Ibarra J. Maize domestication and gene interaction. New Phytologist. 2018;220(2):395–408. doi:https://doi.org/10.1111/nph.15350.

48. Doebley J. The Genetics of Maize Evolution. Annual Review of Genetics. 2004;38(1):37–59. doi:10.1146/annurev.genet.38.072902.092425.

49. Fukunaga K, Hill J, Vigouroux Y, Matsuoka Y, Sanchez G J, Liu K, et al. Genetic diversity and population structure of teosinte. Genetics. 2005;169(4):2241–2254.

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

51. O'Brien AM, Sawers RJH, Strauss SY, Ross-Ibarra J. Adaptive phenotypic divergence in an annual grass differs across biotic contexts. Evolution. 2019;73(11):2230–2246. doi:https://doi.org/10.1111/evo.13818.

52. Staller JE. High Altitude Maize (Zea Mays L.) Cultivation and Endemism in the Lake Titicaca Basin. Journal of Botany Research. 2016;1(11). doi:High Altitude Maize (Zea Mays L.) Cultivation and Endemism in the Lake Titicaca Basin.

53. Chen Q, Samayoa LF, Yang CJ, Bradbury PJ, Olukolu BA, Neumeyer MA, et al. The genetic architecture of the maize progenitor, teosinte, and how it was altered during maize domestication. PLOS Genetics. 2020;16(5):e1008791. doi:10.1371/journal.pgen.1008791.

54. Yang CJ, Samayoa LF, Bradbury PJ, Olukolu BA, Xue W, York AM, et al. The genetic architecture of teosinte catalyzed and constrained maize domestication. Proceedings of the National Academy of Sciences. 2019;116(12):5643–5652. doi:10.1073/pnas.1820997116.

55. van Heerwaarden J, Ross-Ibarra J, Doebley J, Glaubitz JC, Gonzalez JdJS, Gaut BS, et al. Fine scale genetic structure in the wild ancestor of maize (Zea mays ssp. parviglumis). Molecular ecology. 2010;doi:10.1111/j.1365-294X.2010.04559.x.

56. Skotte L, Korneliussen TS, Albrechtsen A. Estimating individual admixture proportions from next generation sequencing data. Genetics. 2013;195(3):693–702.

57. Bradburd GS, Ralph PL, Coop GM. Disentangling the Effects of Geographic and Ecological Isolation on Genetic Differentiation. Evolution. 2013;67(11):3258–3273. doi:10.1111/evo.12193.

58. Corbett-Detig R, Nielsen R. 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. 2017;13(1):e1006529.

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.05.434040; this version posted September 21, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

59. Moorjani P, Patterson N, Hirschhorn JN, Keinan A, Hao L, Atzmon G, et al. The History of African Gene Flow into Southern Europeans, Levantines, and Jews. PLOS Genetics. 2011;7(4):e1001373. doi:10.1371/journal.pgen.1001373.

60. Loh PR, Lipson M, Patterson N, Moorjani P, Pickrell JK, Reich D, et al. Inferring admixture histories of human populations using linkage disequilibrium. Genetics. 2013;193(4):1233–1254.

61. Veller C, Edelman NB, Muralidhar P, Nowak MA. Recombination, variance in genetic relatedness, and selection against introgressed DNA. bioRxiv. 2019. doi:10.1101/846147.

62. Ogut F, Bian Y, Bradbury PJ, Holland JB. Joint-multiple family linkage analysis predicts within-family variation better than single-family analysis of the maize nested association mapping population. Heredity. 2015;114(6):552–563.

63. Schnable PS, Ware D, Fulton RS, Stein JC, Wei F, Pasternak S, et al. The B73 Maize Genome: Complexity, Diversity, and Dynamics. Science. 2009;326(5956):1112–1115. doi:10.1126/science.1178534.

64. Le Corre V, Siol M, Vigouroux Y, Tenaillon MI, Délye C. Adaptive introgression from maize has facilitated the establishment of teosinte as a noxious weed in Europe. Proceedings of the National Academy of Sciences. 2020;117(41):25618–25627. doi:10.1073/pnas.2006633117.

65. Moose SP, Lauter N, Carlson SR. The Maize macrohairless1 Locus Specifically Promotes Leaf Blade Macrohair Initiation and Responds to Factors Regulating Leaf Identity. Genetics. 2004;166(3):1451–1461. doi:10.1534/genetics.166.3.1451.

66. Jiao Y, Peluso P, Shi J, Liang T, Stitzer MC, Wang B, et al. Improved maize reference genome with single-molecule technologies. Nature. 2017;546(76597659):524–527. doi:10.1038/nature22971.

67. Crow T, Ta J, Nojoomi S, Aguilar-Rangel MR, Rodríguez JVT, Gates D, et al. Gene regulatory effects of a large chromosomal inversion in highland maize. PLOS Genetics. 2020;16(12):e1009213. doi:10.1371/journal.pgen.1009213.

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

69. Doebley J, Stec A, Hubbard L. The evolution of apical dominance in maize. Nature. 1997;386(66246624):485–488. doi:10.1038/386485a0.

70. Dong Z, Xiao Y, Govindarajulu R, Feil R, Siddoway ML, Nielsen T, et al. The regulatory landscape of a core maize domestication module controlling bud dormancy and growth repression. Nature Communications. 2019;10(11):3810. doi:10.1038/s41467-019-11774-w.

71. Dorweiler J, Stec A, Kermicle J, Doebley J. Teosinte glume architecture 1: A Genetic Locus Controlling a Key Step in Maize Evolution. Science. 1993;262(5131):233–235. doi:10.1126/science.262.5131.233.

72. Wang H, Nussbaum-Wagler T, Li B, Zhao Q, Vigouroux Y, Faller M, et al. The origin of the naked grains of maize. Nature. 2005;436(70517051):714–719. doi:10.1038/nature03863.

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.05.434040; this version posted September 21, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

73. Wills DM, Fang Z, York AM, Holland JB, Doebley JF. Defining the Role of the MADS-Box Gene, Zea Agamous-like1, a Target of Selection During Maize Domestication. Journal of Heredity. 2018;109(3):333–338. doi:10.1093/jhered/esx073.

74. Wills DM, Whipple CJ, Takuno S, Kursel LE, Shannon LM, Ross-Ibarra J, et al. From Many, One: Genetic Control of Prolificacy during Maize Domestication. PLoS Genetics. 2013;9(6). doi:10.1371/journal.pgen.1003604.

75. Lin Z, Li X, Shannon LM, Yeh CT, Wang ML, Bai G, et al. Parallel domestication of the Shattering1 genes in cereals. Nature Genetics. 2012;44(6):720–724. doi:10.1038/ng.2281.

76. Whitt SR, Wilson LM, Tenaillon MI, Gaut BS, Buckler ES. Genetic diversity and selection in the maize starch pathway. Proceedings of the National Academy of Sciences. 2002;99(20):12959–12962. doi:10.1073/pnas.202476999.

77. Sosso D, Luo D, Li QB, Sasse J, Yang J, Gendrot G, et al. Seed filling in domesticated maize and rice depends on SWEET-mediated hexose transport. Nature Genetics. 2015;47(1212):1489–1493. doi:10.1038/ng.3422.

78. Trimble L, Shuler S, Tracy WF. Characterization of Five Naturally Occurring Alleles at the Sugary1 Locus for Seed Composition, Seedling Emergence, and Isoamylase1 Activity. Crop Science. 2016;56(4):1927–1939. doi:https://doi.org/10.2135/cropsci2015.02.0117.

79. Doebley J, Stec A. Genetic Analysis of the Morphological Differences between Maize and Teosinte. Genetics. 1991;129(1):285–295.

80. Doebley J, Stec A. Inheritance of the morphological differences between maize and teosinte: comparison of results for two F2 populations. Genetics. 1993;134(2):559–570. doi:10.1093/genetics/134.2.559.

81. Bomblies K, Doebley JF. Pleiotropic Effects of the Duplicate Maize FLORICAULA/LEAFY Genes zfl1 and zfl2 on Traits Under Selection During Maize Domestication. Genetics. 2006;172(1):519–531. doi:10.1534/genetics.105.048595.

82. Wang Z, Ueda T, Messing J. Characterization of the maize prolamin box-binding factor-1 (PBF-1) and its role in the developmental regulation of the zein multigene family. Gene. 1998;223(1):321–332. doi:10.1016/S0378-1119(98)00244-3.

83. Gallavotti A, Zhao Q, Kyozuka J, Meeley RB, Ritter MK, Doebley JF, et al. The role of barren stalk1 in the architecture of maize. Nature. 2004;432(70177017):630–635. doi:10.1038/nature03148.

84. Vollbrecht E, Springer PS, Goh L, Buckler Iv ES, Martienssen R. Architecture of floral branch systems in maize and related grasses. Nature. 2005;436(70547054):1119–1126. doi:10.1038/nature03892.

85. Sigmon B, Vollbrecht E. Evidence of selection at the ramosa1 locus during maize domestication. Molecular Ecology. 2010;19(7):1296–1311. doi:https://doi.org/10.1111/j.1365-294X.2010.04562.x.

86. Rodríguez-Zapata F, Barnes AC, Blöcher-Juárez KA, Gates D, Kur A, Wang L, et al. Teosinte introgression modulates phosphatidylcholine levels and induces early maize flowering time. bioRxiv. 2021;doi:10.1101/2021.01.25.426574.

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.05.434040; this version posted September 21, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

87. Dong Z, Danilevskaya O, Abadie T, Messina C, Coles N, Cooper M. A Gene Regulatory Network Model for Floral Transition of the Shoot Apex in Maize and Its Dynamic Modeling. PLOS ONE. 2012;7(8):e43450. doi:10.1371/journal.pone.0043450.

88. Li Yx, Li C, Bradbury PJ, Liu X, Lu F, Romay CM, et al. Identification of genetic variants associated with maize flowering time using an extremely large multi-genetic background population. The Plant Journal. 2016;86(5):391–402. doi:https://doi.org/10.1111/tpj.13174.

89. Buckler ES, Holland JB, Bradbury PJ, Acharya CB, Brown PJ, Browne C, et al. The genetic architecture of maize flowering time. Science. 2009;325(5941):714–718. doi:10.1126/science.1174276.

90. Rowan BA, Heavens D, Feuerborn TR, Tock AJ, Henderson IR, Weigel D. An ultra high-density Arabidopsis thaliana crossover map that refines the influences of structural variation and epigenetic features. Genetics. 2019;213(3):771–787.

91. R Core Team. R: A Language and Environment for Statistical Computing; 2019. Available from: `https://www.R-project.org/`.

92. Wingett SW, Andrews S. FastQ Screen: A tool for multi-genome mapping and quality control. F1000Research. 2018;7:1338. doi:10.12688/f1000research.15931.2.

93. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30(15):2114–2120. doi:10.1093/bioinformatics/btu170.

94. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv [q-bio]. 2013;.

95. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009;25(16):2078–2079. doi:10.1093/bioinformatics/btp352.

96. Li H. Improving SNP discovery by base alignment quality. Bioinformatics. 2011;27(8):1157–1158. doi:10.1093/bioinformatics/btr076.

97. Korneliussen TS, Albrechtsen A, Nielsen R. ANGSD: Analysis of Next Generation Sequencing Data. BMC bioinformatics. 2014;15(1):356.

98. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 2010;26(6):841–842. doi:10.1093/bioinformatics/btq033.

99. Tenaillon MI, Sawkins MC, Long AD, Gaut RL, Doebley JF, Gaut BS. Patterns of DNA sequence polymorphism along chromosome 1 of maize (Zea mays ssp. mays L.). Proceedings of the National Academy of Sciences. 2001;98(16):9161–9166. doi:10.1073/pnas.151244298.

100. Meisner J, Albrechtsen A. Inferring Population Structure and Admixture Proportions in Low-Depth NGS Data. Genetics. 2018;210(2):719–731. doi:10.1534/genetics.118.301336.

101. Hudson RR, Slatkin M, Maddison WP. Estimation of levels of gene flow from DNA sequence data. Genetics. 1992;132(2):583–589.

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.05.434040; this version posted September 21, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

102. Bhatia G, Patterson N, Sankararaman S, Price AL. Estimating and interpreting FST: The impact of rare variants. Genome Research. 2013;23(9):1514–1521. doi:10.1101/gr.154831.113.

103. Yi X, Liang Y, Huerta-Sanchez E, Jin X, Cuo ZXP, Pool JE, et al. Sequencing of Fifty Human Exomes Reveals Adaptation to High Altitude. Science (New York, NY). 2010;329(5987):75–78. doi:10.1126/science.1190371.

104. Haider S, Waggott D, C Boutros P. bedr: Genomic Region Processing using Tools Such as 'BEDTools', 'BEDOPS' and 'Tabix'; 2019. Available from: `https://CRAN.R-project.org/package=bedr`.

105. Green RE, Krause J, Briggs AW, Maricic T, Stenzel U, Kircher M, et al. A Draft Sequence of the Neandertal Genome. Science. 2010;328(5979):710–722. doi:10.1126/science.1188021.

106. Peter BM. Admixture, Population Structure, and F-Statistics. Genetics. 2016;202(4):1485–1501. doi:10.1534/genetics.115.183913.

107. Soraggi S, Wiuf C, Albrechtsen A. Powerful Inference with the D-Statistic on Low-Coverage Whole-Genome Data. G3: Genes—Genomes—Genetics. 2018;8(2):551–566. doi:10.1534/g3.117.300192.

108. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society: Series B (Methodological). 1995;57(1):289–300. doi:https://doi.org/10.1111/j.2517-6161.1995.tb02031.x.

109. Portwood I John L, Woodhouse MR, Cannon EK, Gardiner JM, Harper LC, Schaeffer ML, et al. MaizeGDB 2018: the maize multi-genome genetics and genomics database. Nucleic Acids Research. 2019;47(D1):D1146–D1154. doi:10.1093/nar/gky1046.

110. Köster J, Rahmann S. Snakemake—a scalable bioinformatics workflow engine. Bioinformatics. 2012;28(19):2520–2522. doi:10.1093/bioinformatics/bts480.

111. Wickham H, Averick M, Bryan J, Chang W, McGowan LD, François R, et al. Welcome to the tidyverse. Journal of Open Source Software. 2019;4(43):1686. doi:10.21105/joss.01686.

112. Garnier S. viridis: Default Color Maps from 'matplotlib'; 2018. Available from: `https://CRAN.R-project.org/package=viridis`.

113. Canty A, Ripley BD. boot: Bootstrap R (S-Plus) Functions; 2019.

114. Davison AC, Hinkley DV. Bootstrap Methods and Their Applications. Cambridge: Cambridge University Press; 1997. Available from: `http://statwww.epfl.ch/davison/BMA/`.

115. Auguie B. gridExtra: Miscellaneous Functions for "Grid" Graphics; 2017. Available from: `https://CRAN.R-project.org/package=gridExtra`.

116. Ahlmann-Eltze C. ggupset: Combination Matrix Axis for 'ggplot2' to Create 'UpSet' Plots; 2020. Available from: `https://CRAN.R-project.org/package=ggupset`.

117. Pedersen TL. tidygraph: A Tidy API for Graph Manipulation; 2020. Available from: `https://CRAN.R-project.org/package=tidygraph`.

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.05.434040; this version posted September 21, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Supporting Information

S1 Table. Population metadata.

S2 Table. Parviglumis SRA IDs.

S3 Table. Spearman's rank correlation between genomewide admixture proportions (NGSAdmix) and recombination rate (or coding bp per cM) quintiles

S4 Table. *Mexicana* 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.

S5 Table. Spearman's rank correlation between mean 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.

S6 Table. Diversity within maize and *mexicana* alleles at *inv9f*. Pairwise diversity (π) and Watterson's theta (θ_W) for samples clustering by PCA with the maize-allele at the inversion (PC1 $>$ 0.2) or *mexicana*-allele at the inversion (PC1 $<$ 0.4). Individuals heterozygous for the inversion (or ambiguous in PCA clustering) were excluded. Diversity estimates within the putative inversion region were calculated for each group using the same ANGSD/realSFS pipeline as genomewide diversity estimates. Only subspecies with $>$ 5 samples in a cluster were analysed.

S7 Table. Domestication genes and overlap with introgression deserts.

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.05.434040; this version posted September 21, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

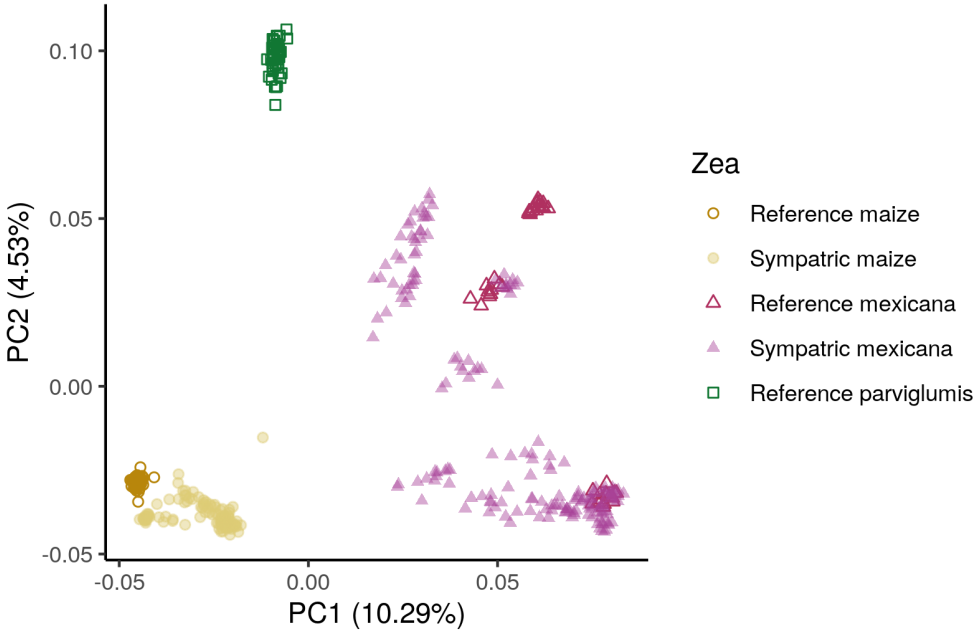

Fig S1. PCA of maize, *mexicana* and *parviglumis*. First and second principal components from the genomewide genetic covariance matrix relating all maize, *mexicana* and *parviglumis* individuals. PC1 aligns with a maize to *mexicana* ancestry gradient while PC2 separates out *parviglumis* ancestry from the other two *Zea* subspecies.

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.05.434040; this version posted September 21, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

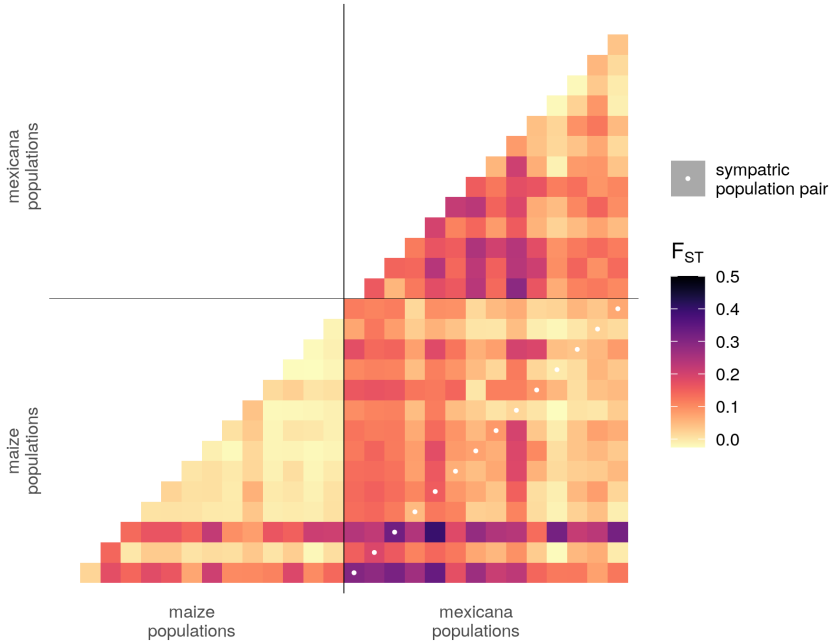

Fig S2. F_{ST} **between *parviglumis* ancestry tracts from different populations** Pairwise F_{ST} between parviglumis ancestry tracts from population 1 (x-axis) and population 2 (y-axis). Populations are sorted by subspecies, then elevation. Local sympatric maize-*mexicana* population pairs are highlighted with a white dot and do not show reduced F_{ST} within *parviglumis* ancestry relative to other (non-local) maize-*mexicana* comparisons.

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.05.434040; this version posted September 21, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

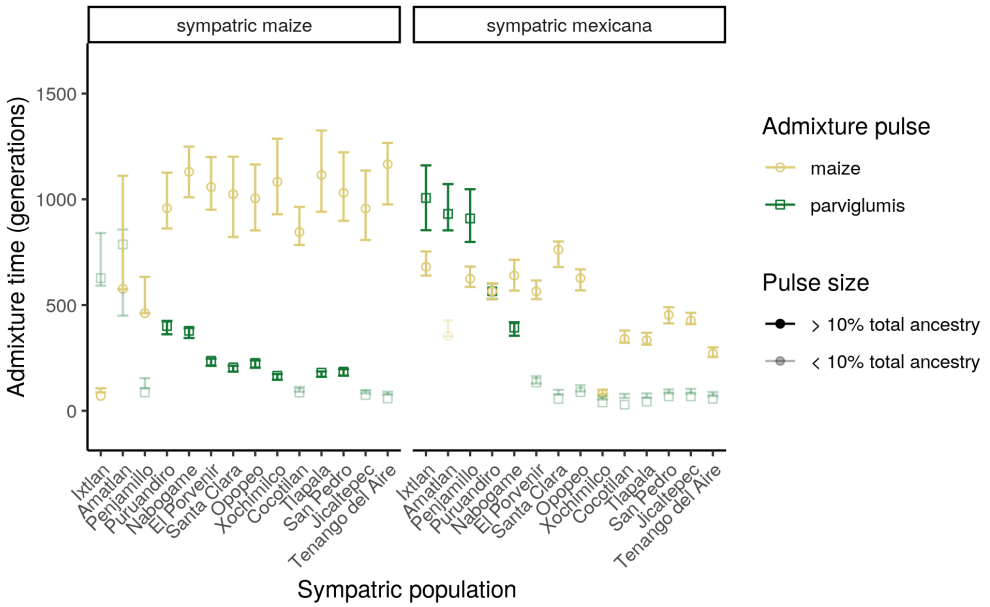

Fig S3. Time since admixture. Estimated generations since admixture under a three-way admixture model: a founding *mexicana* population receives a pulse of *parviglumis* admixture and then a second pulse of maize admixture (possibly resulting in majority maize ancestry). Each sympatric maize and *mexicana* population was analyzed separately under this model, with 95% percentile confidence intervals for each admixture pulse 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. Populations with very small genetic contributions from an ancestry pulse ($< 10\%$) are faded because their timing estimates are less certain.

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.05.434040; this version posted September 21, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

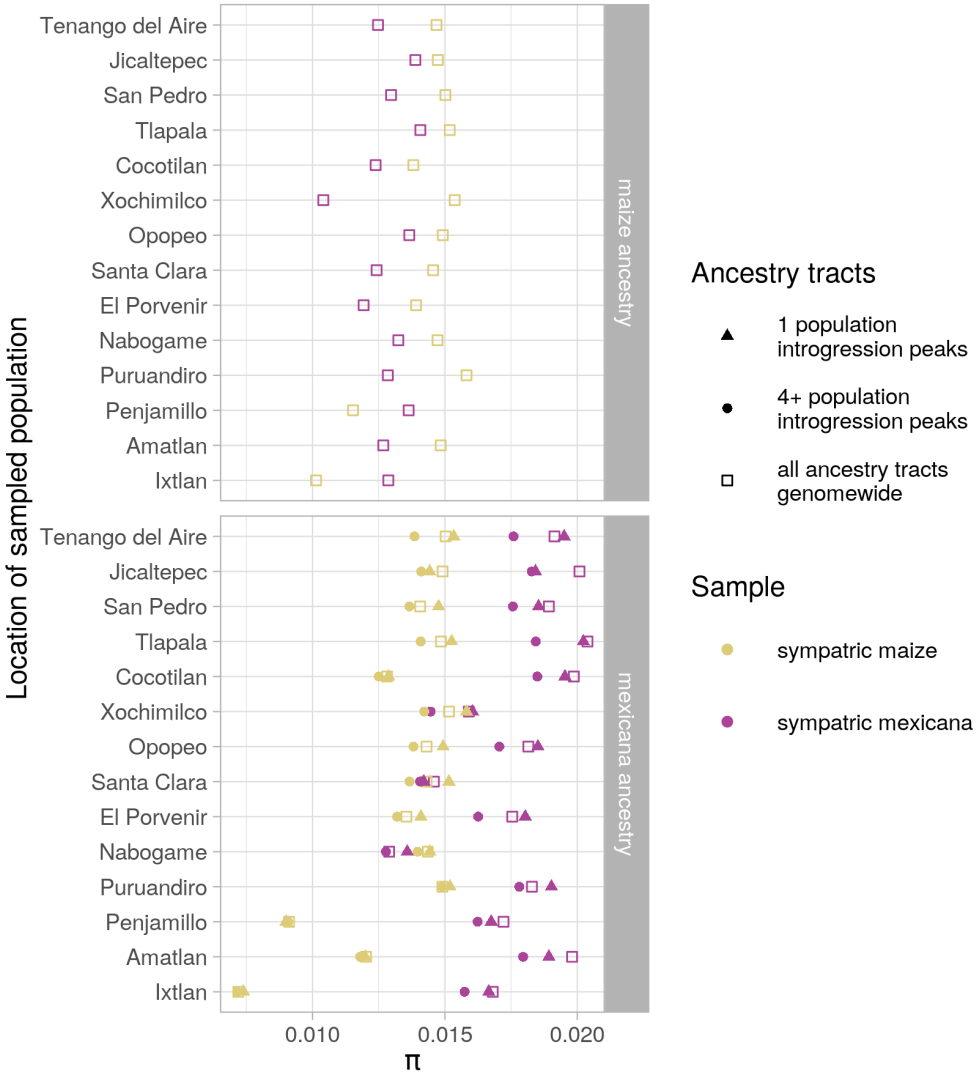

Fig S4. Diversity (π) within ancestry Each point summarises pairwise genetic diversity (π) for genomic regions with high-confidence homozygous maize or *mexicana* ancestry, calculated separately for the sympatric maize and *mexicana* populations at each sampled location. For maize ancestry (top), only a genomewide π is estimated, using all regions with high-confidence homozygous maize ancestry. For *mexicana* ancestry (bottom), 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, introgression peaks shared between the focal maize and at least 3 other maize populations, and a genomewide estimate.

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.05.434040; this version posted September 21, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

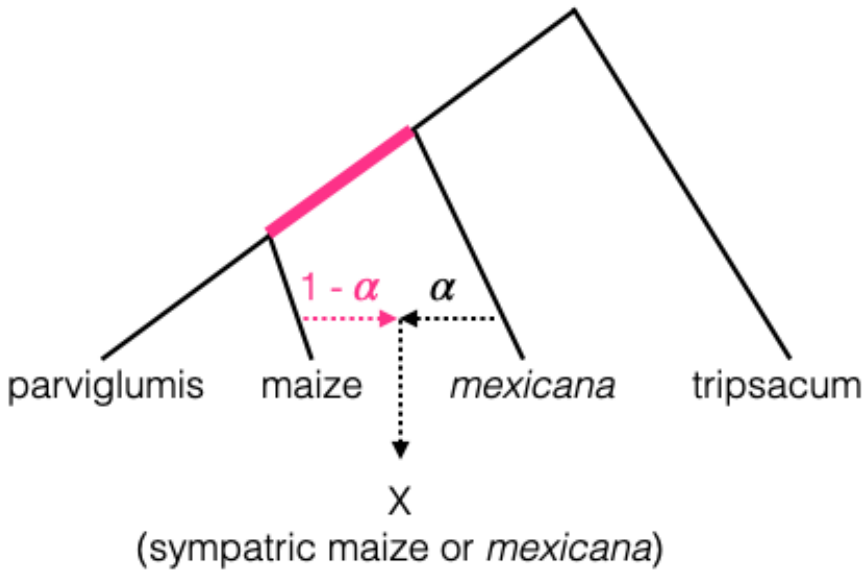

Fig S5. Population tree Phylogenetic tree assumed when estimating the ratio of f_4 statistics. The pink branch represents the shared drift between maize and *parviglumis* that is introduced to the focal sympatric population via admixture of proportion $1 - \alpha$. We used only plants from the Amecameca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

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.05.434040; this version posted September 21, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

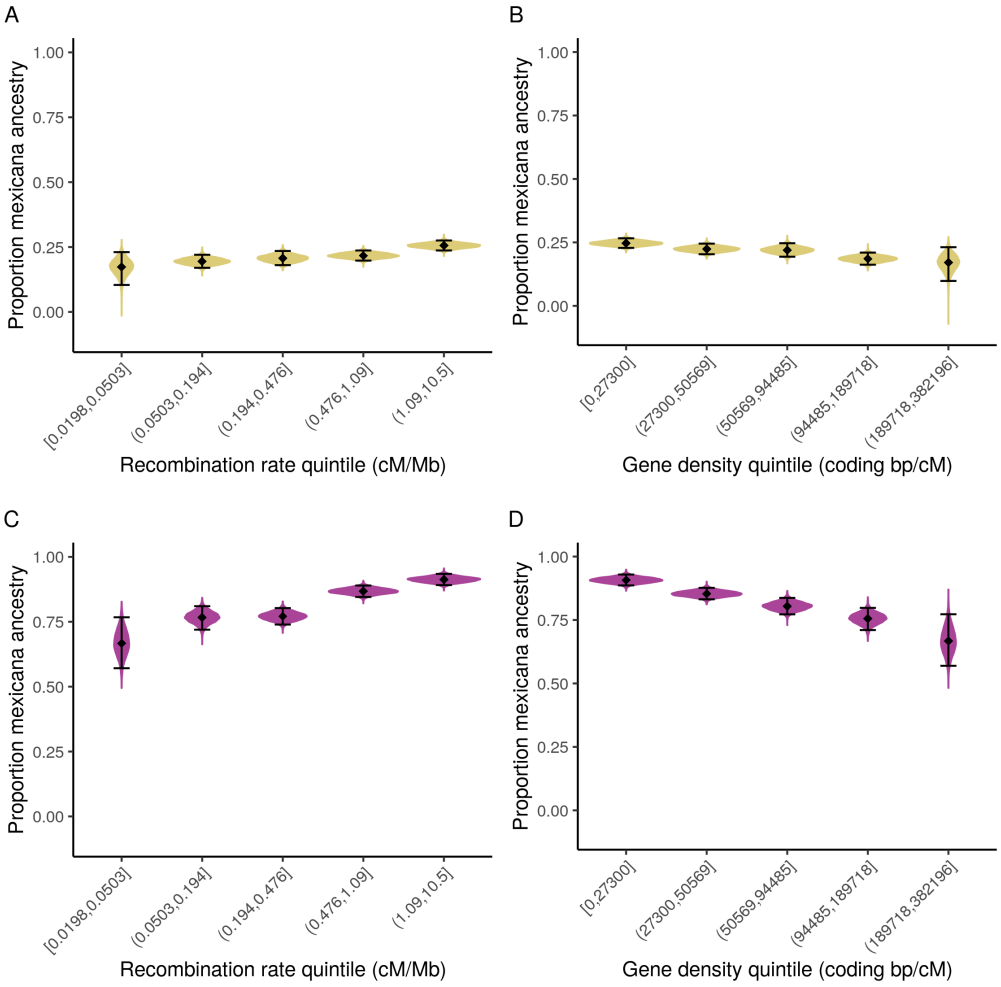

Fig S6. f_4 **ancestry by recombination rate and gene density**. *Mexicana* ancestry proportions by genomic quintiles in sympatric maize (top) and *mexicana* (bottom), estimated using f_4 ratios. Spearman's rank correlations for each plot: (A) *Mexicana* ancestry by recombination rate quintile in sympatric maize ($\rho = 1.00$, $\text{CI}_{95}[0.30,\ 1.00]$), (B) *mexicana* ancestry by gene density (coding bp/cM) quintile in symaptric maize ($\rho = -1.00$, $\text{CI}_{95}[-1.00,\ -0.40]$), (C) *Mexicana* ancestry by recombination rate quintile in sympatric *mexicana* ($\rho = 1.00$, $\text{CI}_{95}[0.70,\ 1.00]$), and (D) *mexicana* ancestry by gene density (coding bp/cM) quintile in symaptric *mexicana* ($\rho = -1.00$, $\text{CI}_{95}[-1.00,\ -0.90]$). Mean ancestry per 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. Note: Ancestry estimates from f_4's are less reliable for sympatric *mexicana* than sympatric maize because of significant *parviglumis* ancestry in some sympatric *mexicana* populations. The f_4 ratio estimation method assumes no additional unmodeled admixture on the population tree (see S5 Fig).

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.05.434040; this version posted September 21, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

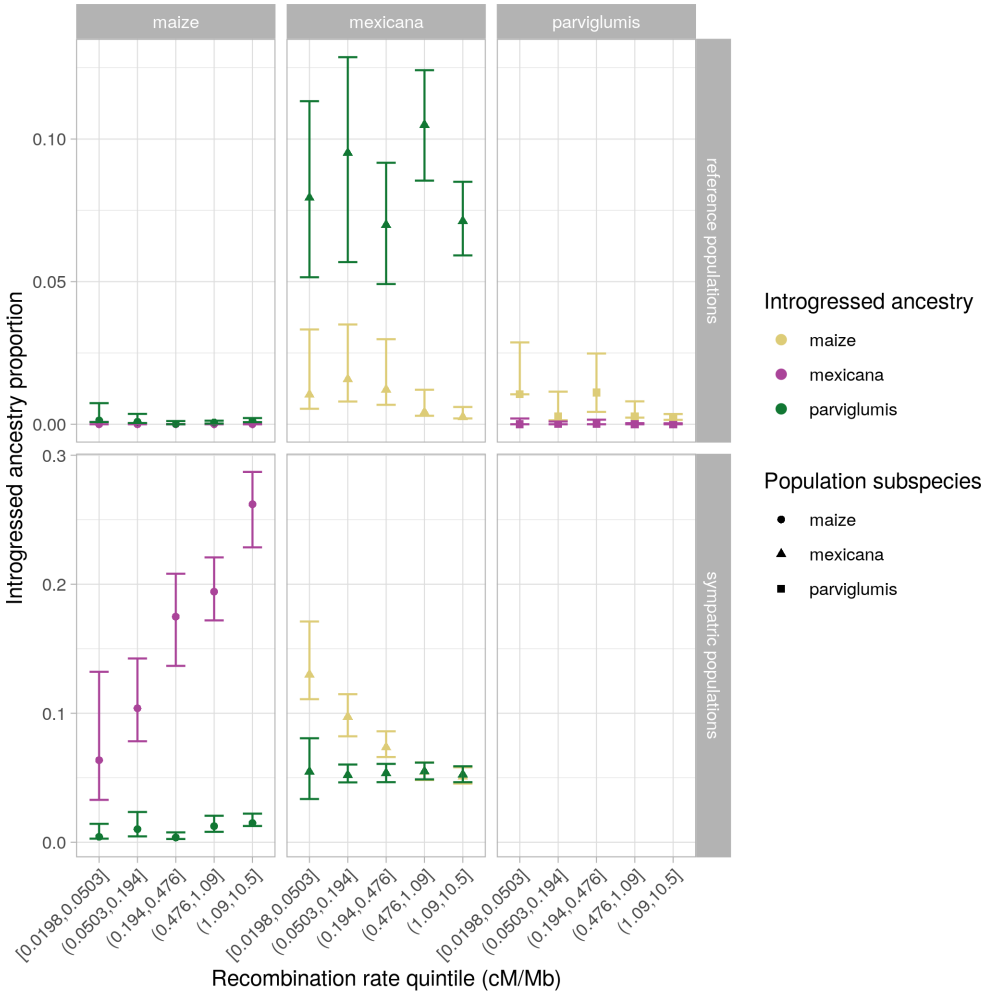

Fig S7. Introgressed ancestry by recombination rate. Inferred introgressed ancestry in reference populations (top) and sympatric maize and *mexicana* populations (bottom) using NGSAdmix (K=3) by recombination rate quintiles. Group mean and 95% percentile bootstrap confidence interval (n = 100) are shown. Different colors distinguish the different introgressing ancestries, and different shapes distinguish the *Zea* subspecies for the sampled populations.

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.05.434040; this version posted September 21, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

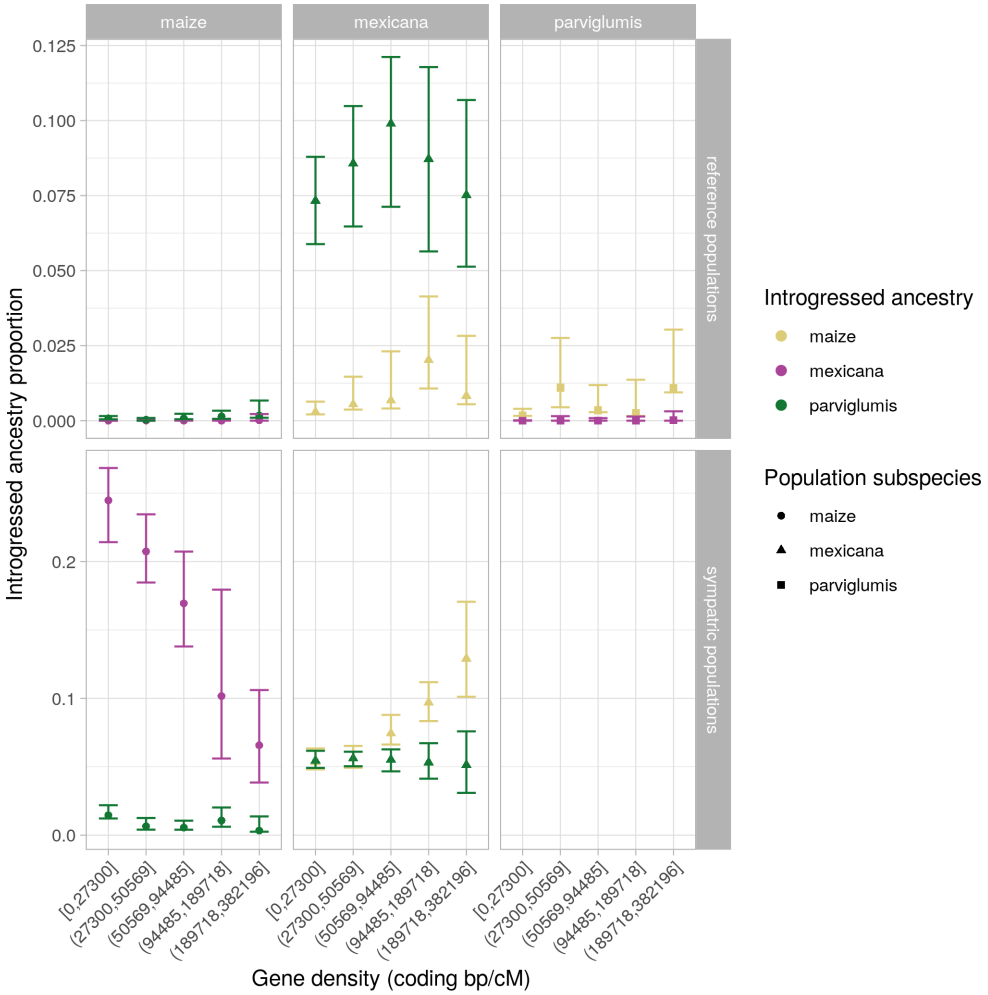

Fig S8. Introgressed ancestry by coding bp per cM. Inferred introgressed ancestry in reference populations (top) and sympatric maize and *mexicana* populations (bottom) using NGSAdmix (K=3) by coding density quintiles. Group mean and 95% percentile bootstrap confidence interval (n = 100) are shown. Different colors distinguish the different introgressing ancestries, and different shapes distinguish the *Zea* subspecies for the sampled populations.

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.05.434040; this version posted September 21, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

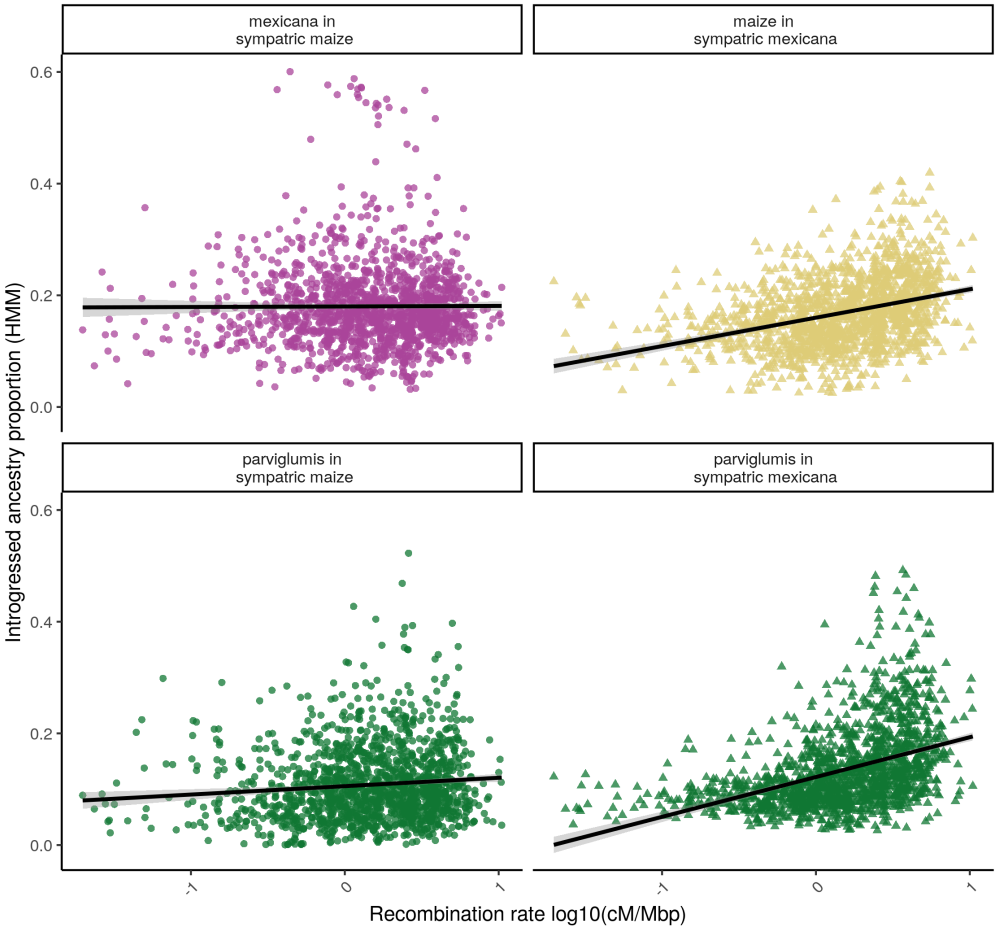

Fig S9. Local introgressed ancestry in 1 cM windows by recombination rate. Estimated local ancestry in sympatric maize and *mexicana* samples using ancestry_hmm. Each point is a 1 cM genomic window and the line shows the best linear model fit for mean introgressed ancestry by recombination rate on a log scale. Different colors distinguish the different introgressing ancestries, and different shapes distinguish the *Zea* subspecies for the sampled sympatric populations.

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.05.434040; this version posted September 21, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

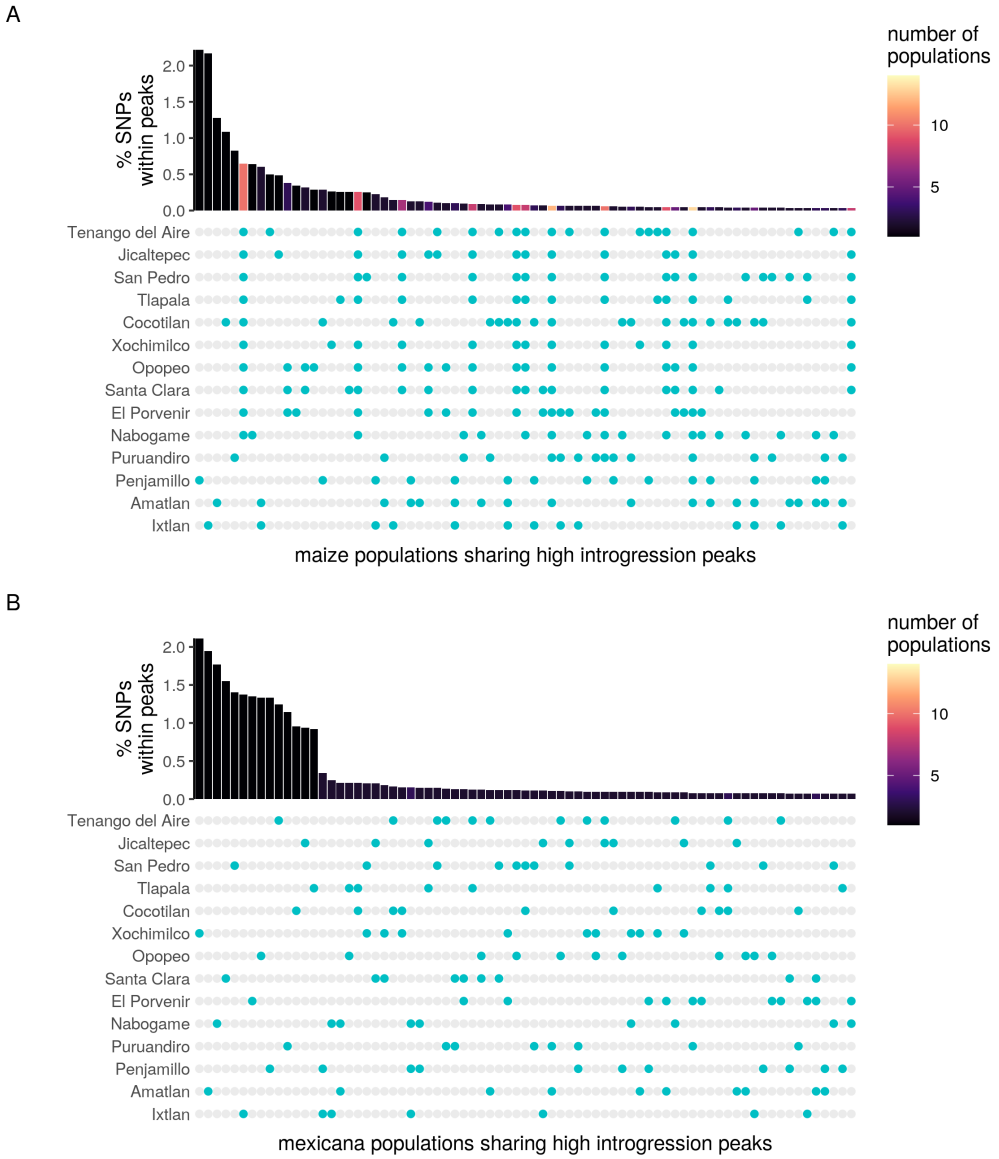

Fig S10. High introgression peaks shared across sympatric populations Here we show the 75 most common combinations of populations that share ancestry peaks (introgressed ancestry > 2 s.d. above each population's mean ancestry) for (A) sympatric maize and (B) sympatric *mexicana*. Bar height represents the percent of SNPs genomewide within peaks shared by the populations highlighted in blue below. Only for sympatric maize do we find that larger sets of populations (brighter colored bars) commonly share peaks across the genome. Populations are ordered from high to low elevation (top to bottom), showing that introgression peak sharing is more common among combinations of the highest elevation maize populations.

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.05.434040; this version posted September 21, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

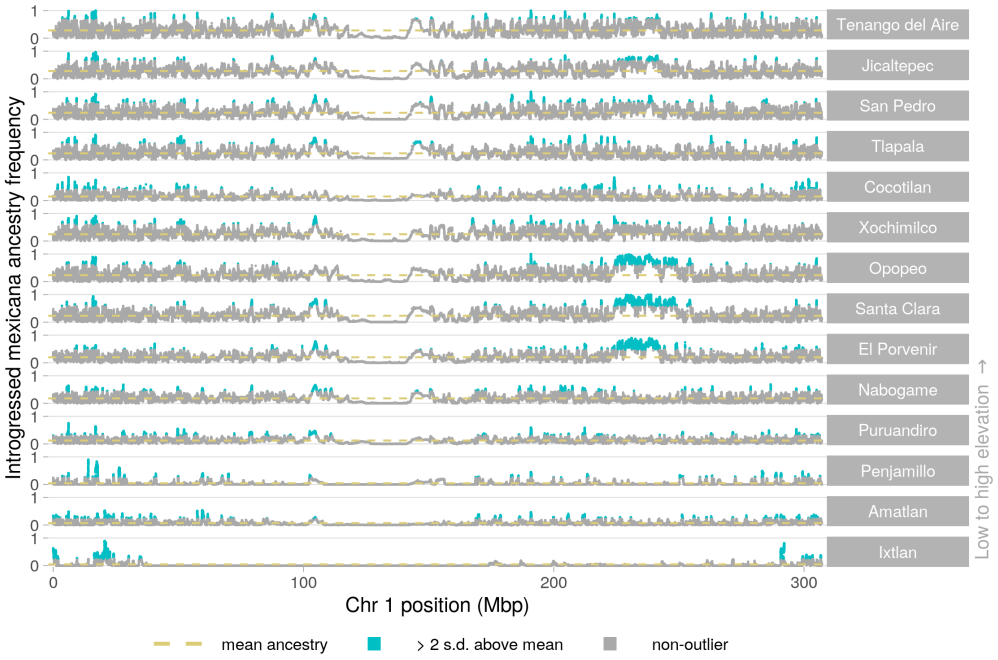

Fig S11. Introgression in maize populations across chromosome 1

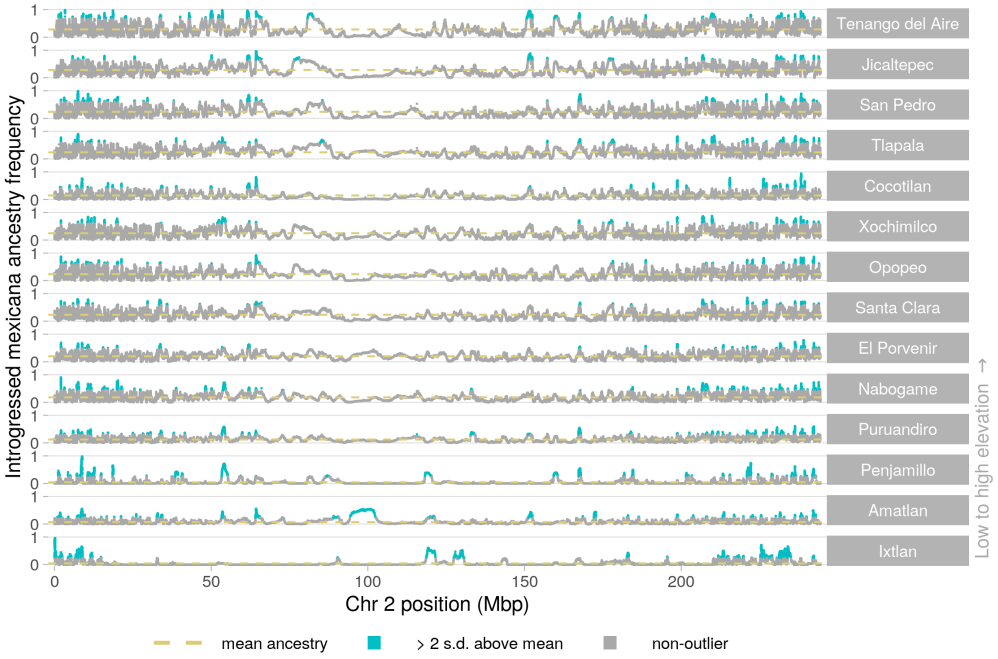

Fig S12. Introgression in maize populations across chromosome 2

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.05.434040; this version posted September 21, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

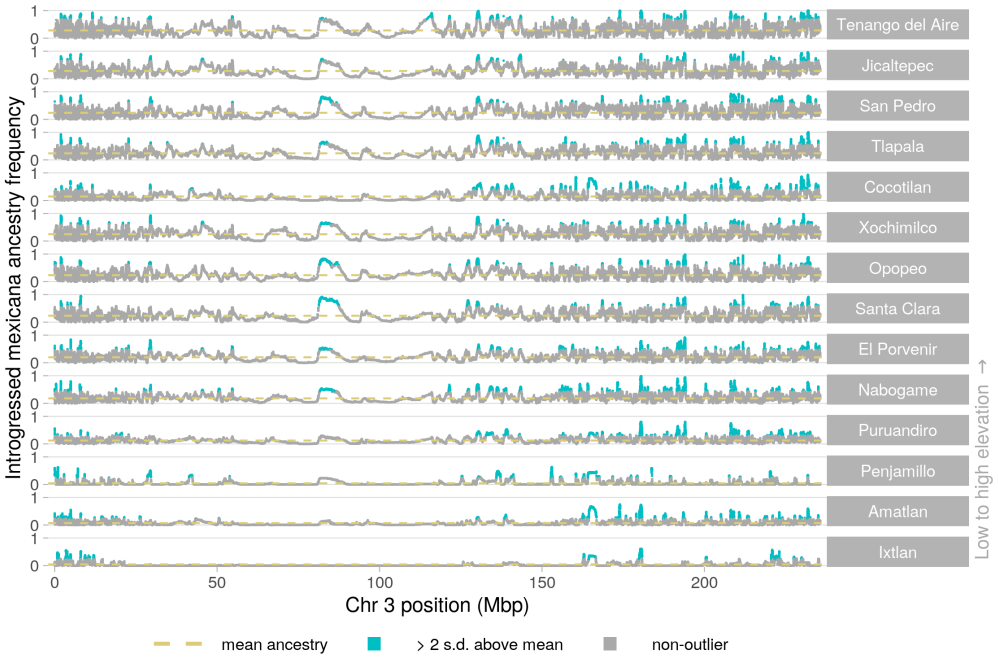

Fig S13. Introgression in maize populations across chromosome 3

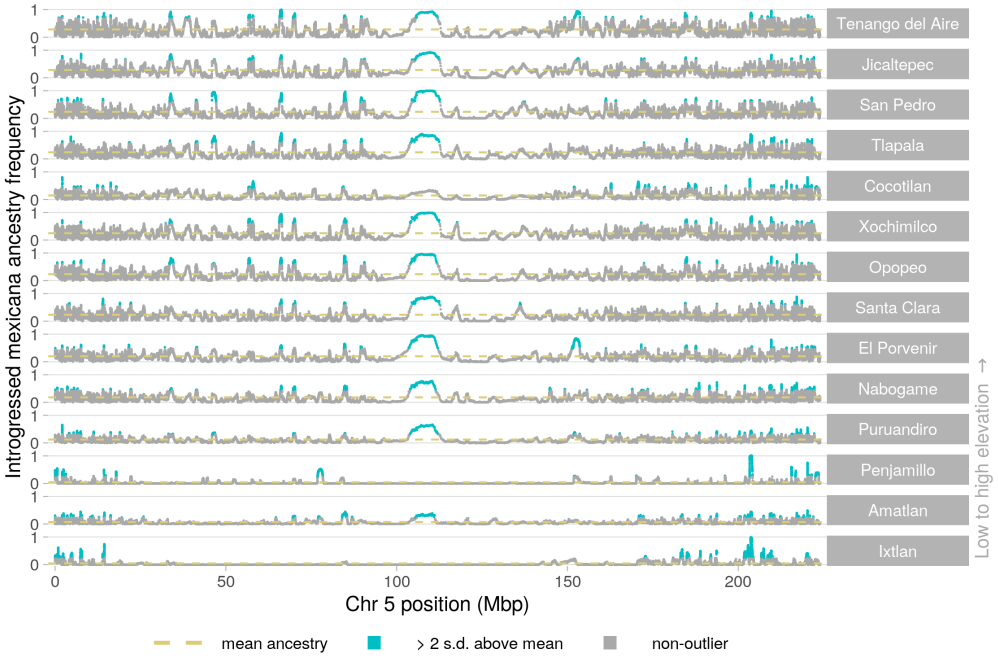

Fig S14. Introgression in maize populations across chromosome 5

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.05.434040; this version posted September 21, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

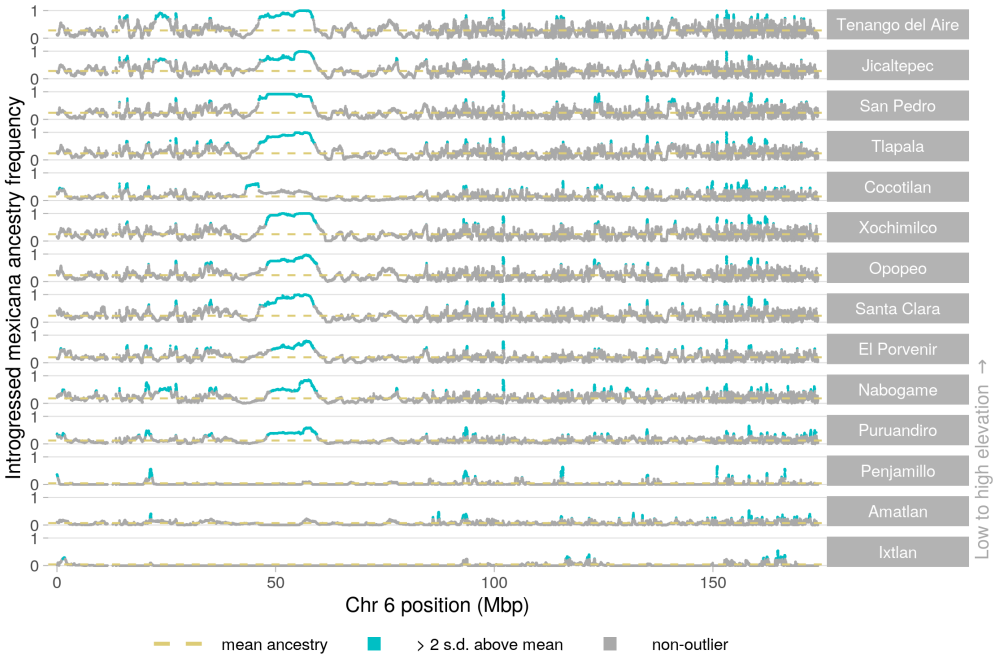

Fig S15. Introgression in maize populations across chromosome 6

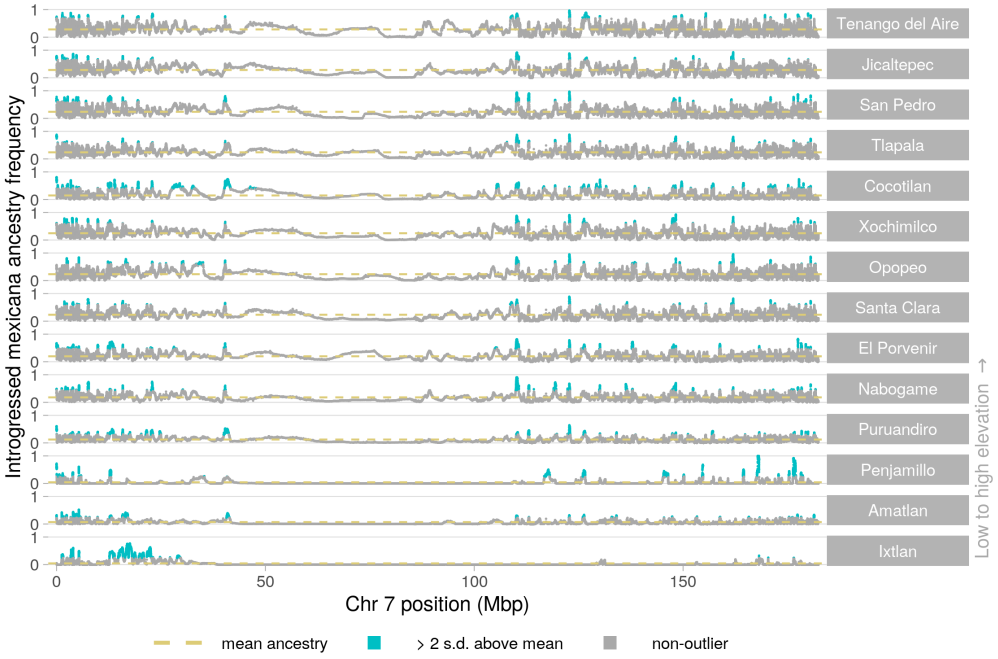

Fig S16. Introgression in maize populations across chromosome 7

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.05.434040; this version posted September 21, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

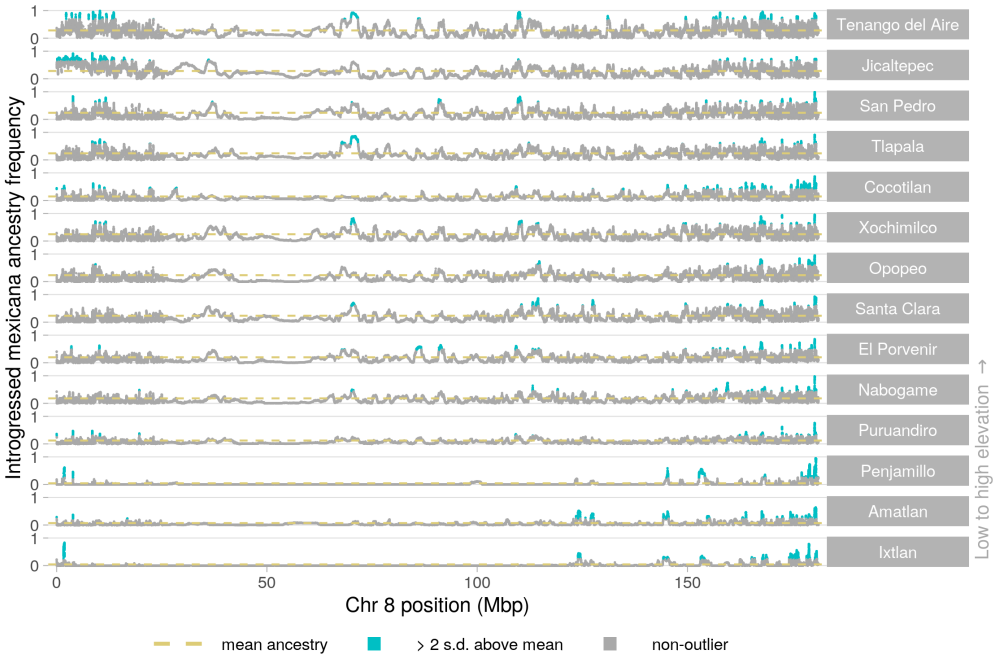

Fig S17. Introgression in maize populations across chromosome 8

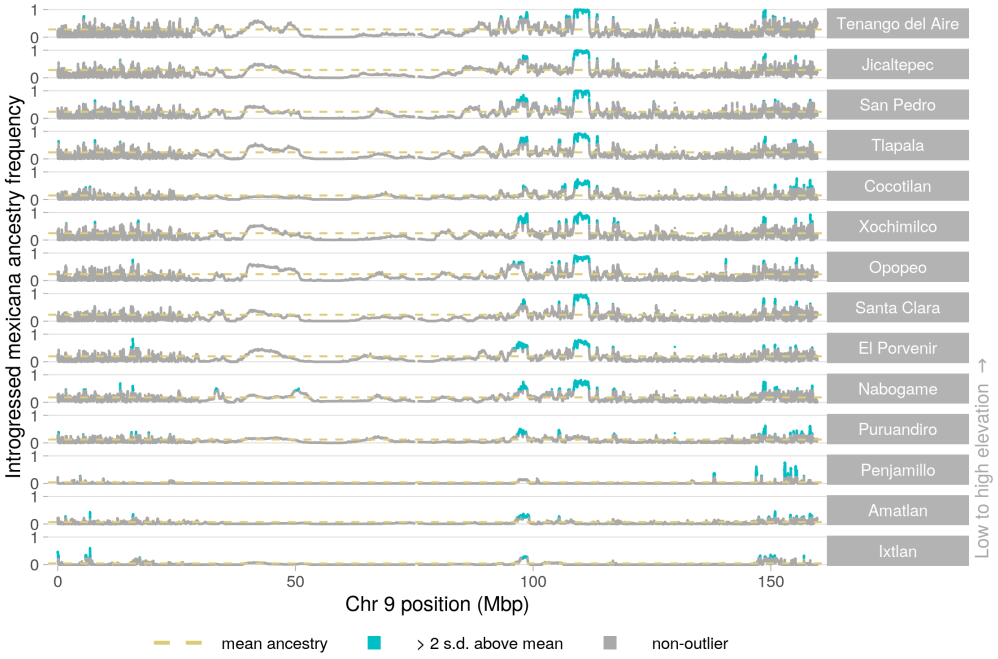

Fig S18. Introgression in maize populations across chromosome 9

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.05.434040; this version posted September 21, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

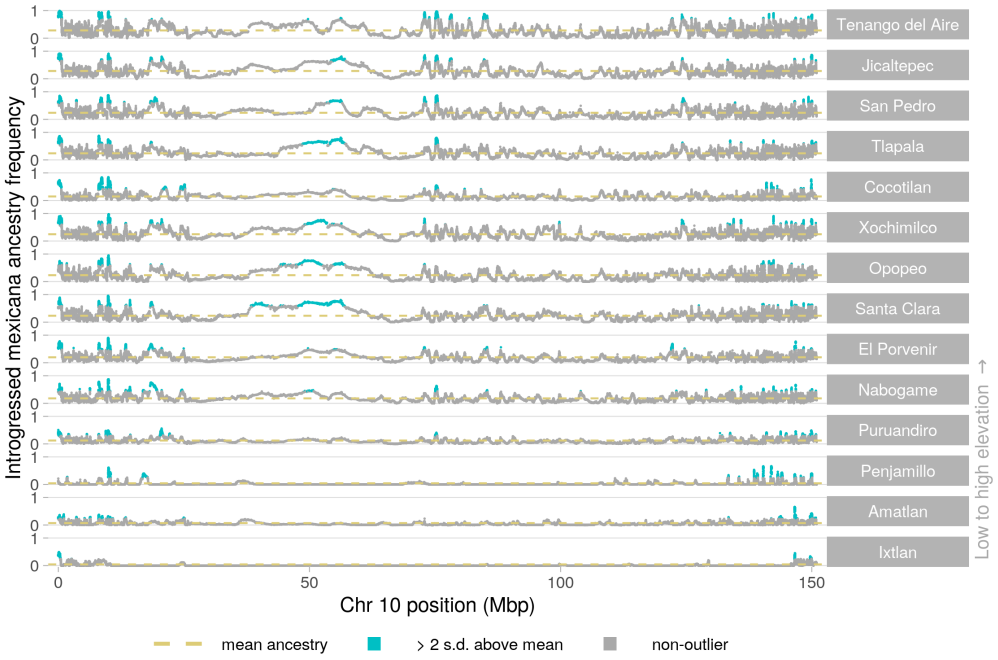

Fig S19. Introgression in maize populations across chromosome 10

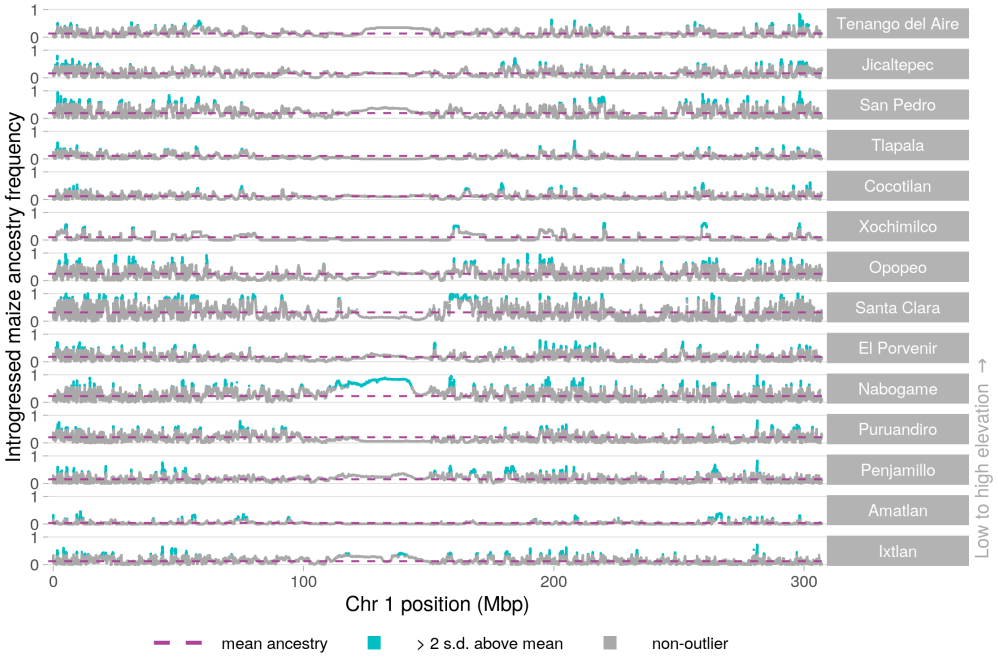

Fig S20. Introgression in *mexicana* populations across chromosome 1

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.05.434040; this version posted September 21, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

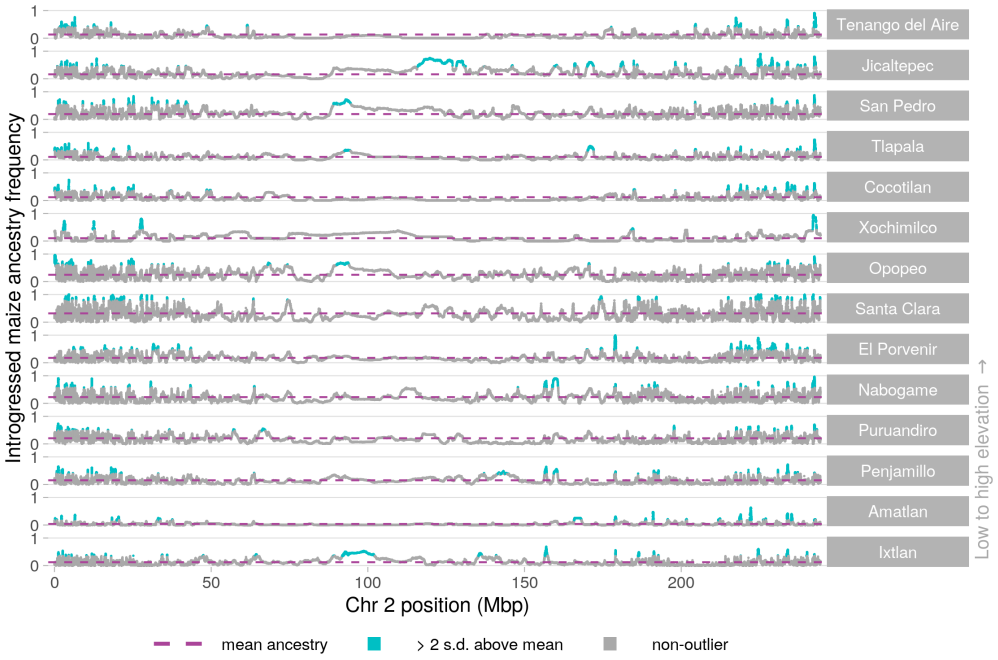

Fig S21. Introgression in *mexicana* populations across chromosome 2

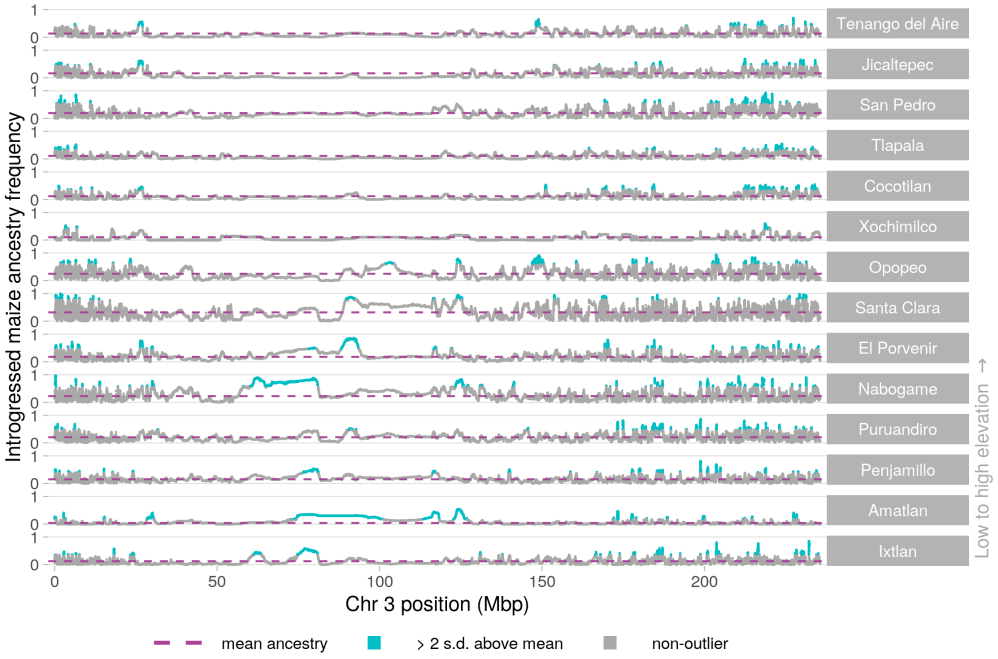

Fig S22. Introgression in *mexicana* populations across chromosome 3

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.05.434040; this version posted September 21, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

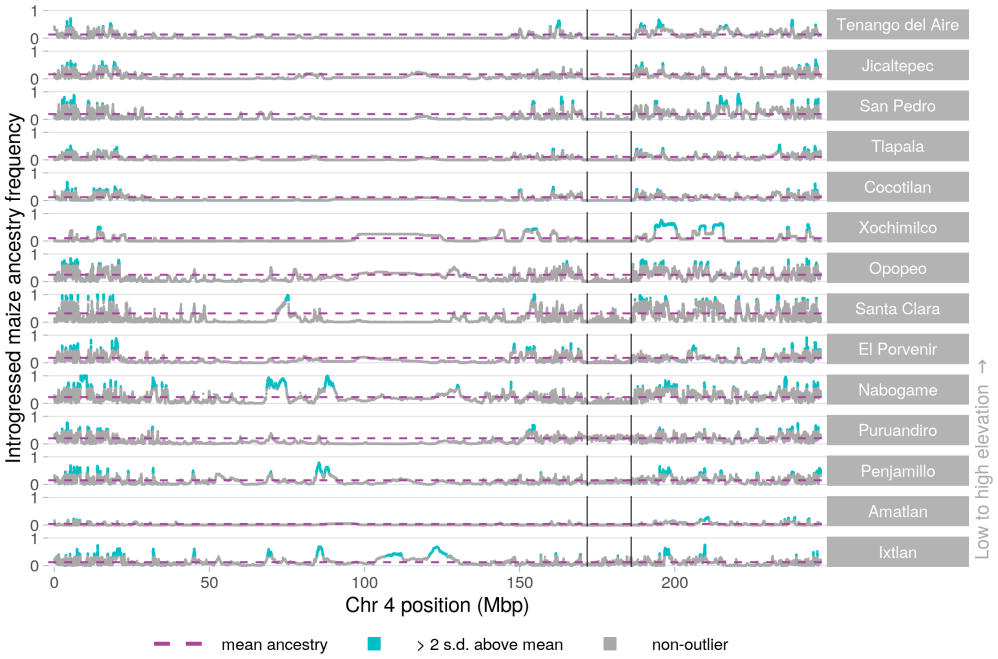

Fig S23. Introgression in *mexicana* populations across chromosome 4
Vertical lines indicate the coordinates for *Inv4m*.

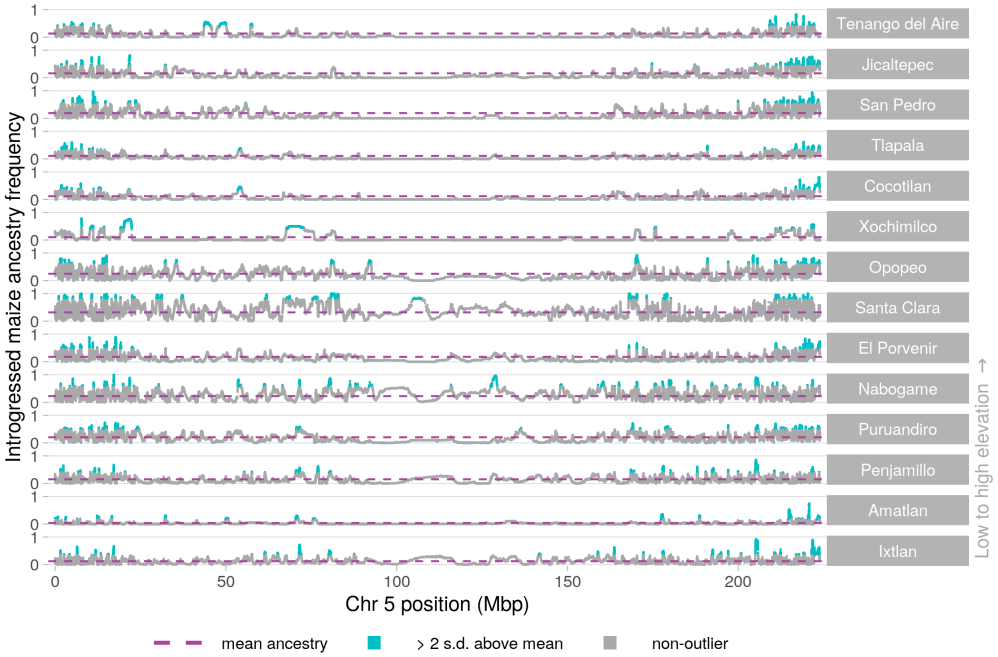

Fig S24. Introgression in *mexicana* populations across chromosome 5

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.05.434040; this version posted September 21, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

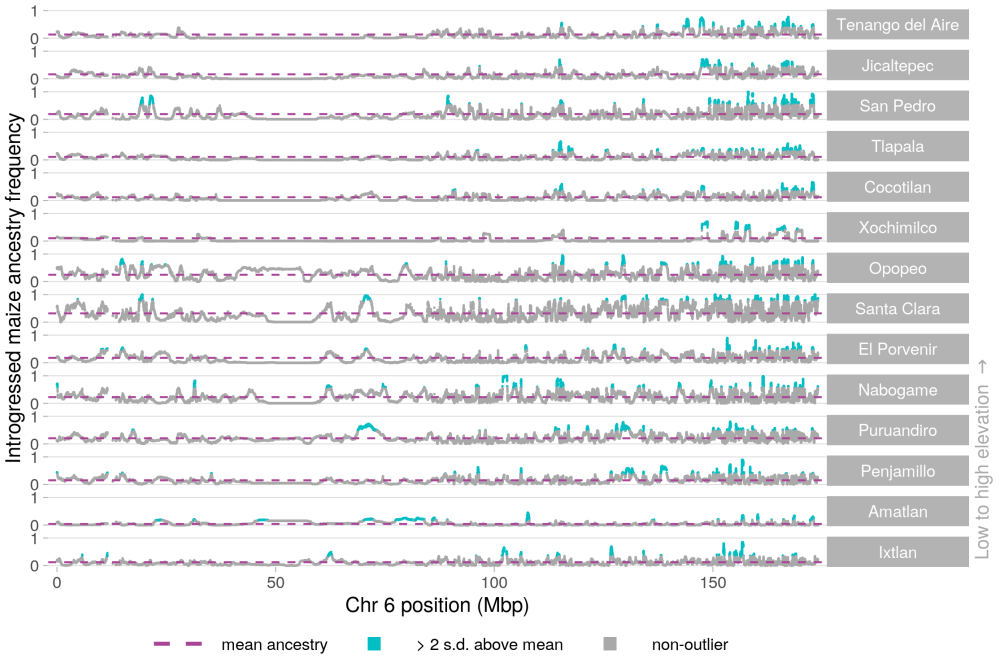

Fig S25. Introgression in *mexicana* populations across chromosome 6

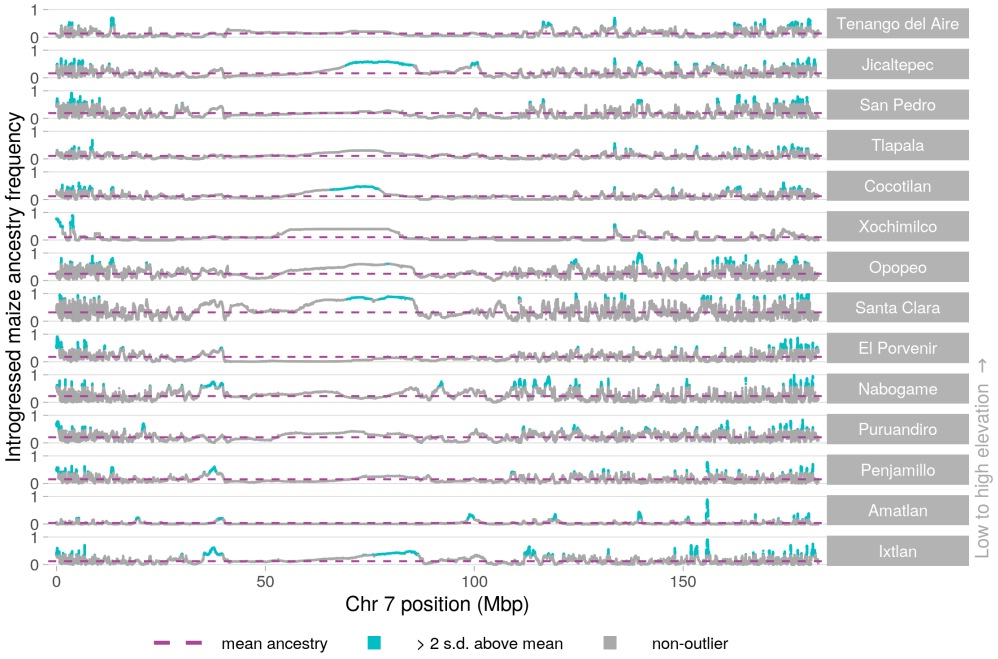

Fig S26. Introgression in *mexicana* populations across chromosome 7

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.05.434040; this version posted September 21, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

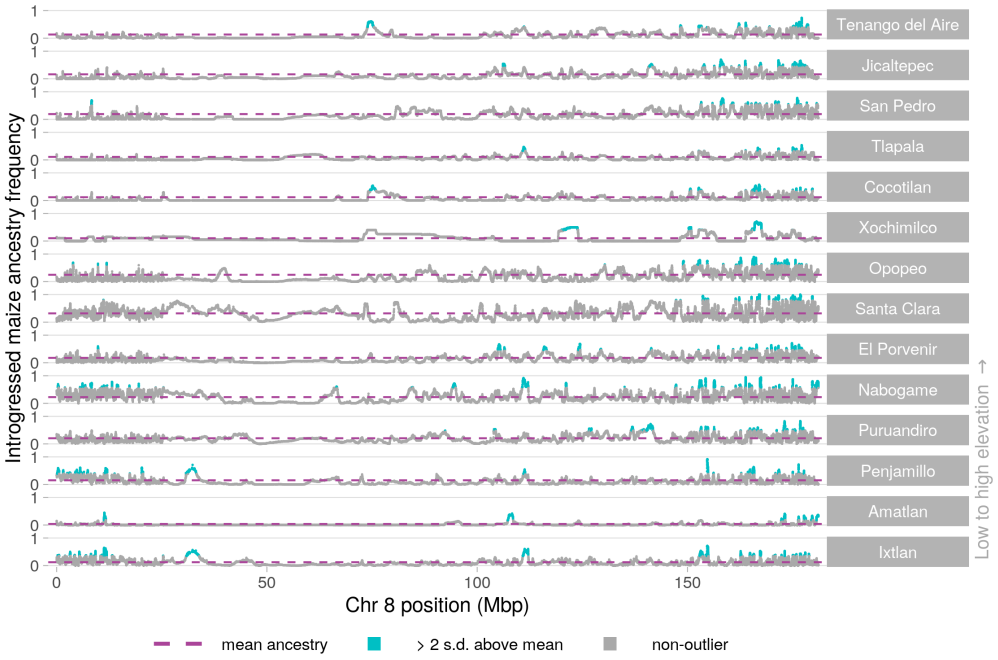

Fig S27. Introgression in *mexicana* populations across chromosome 8

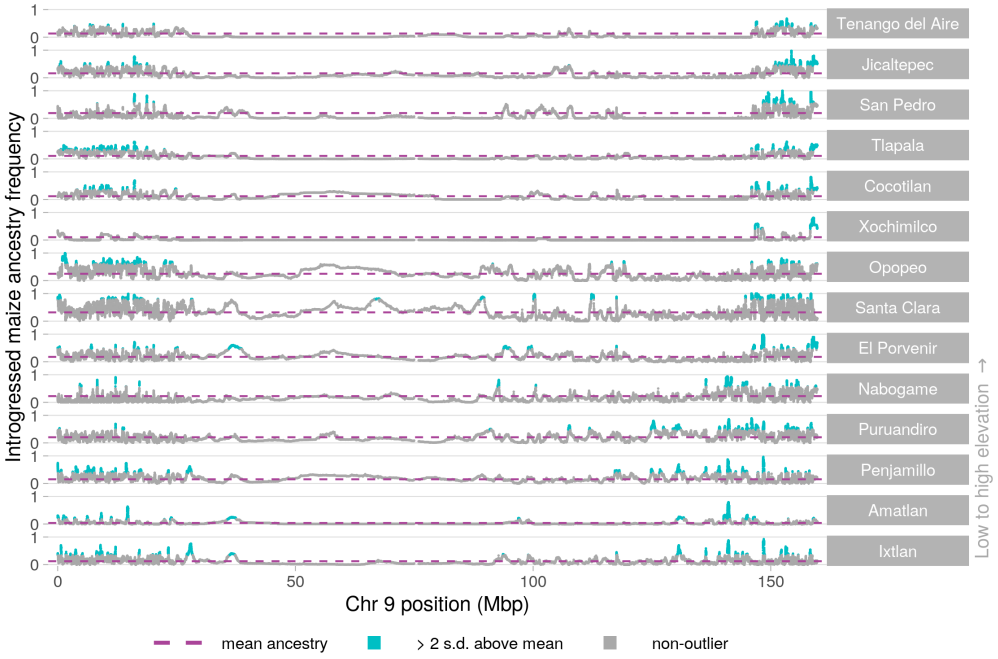

Fig S28. Introgression in *mexicana* populations across chromosome 9

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.05.434040; this version posted September 21, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

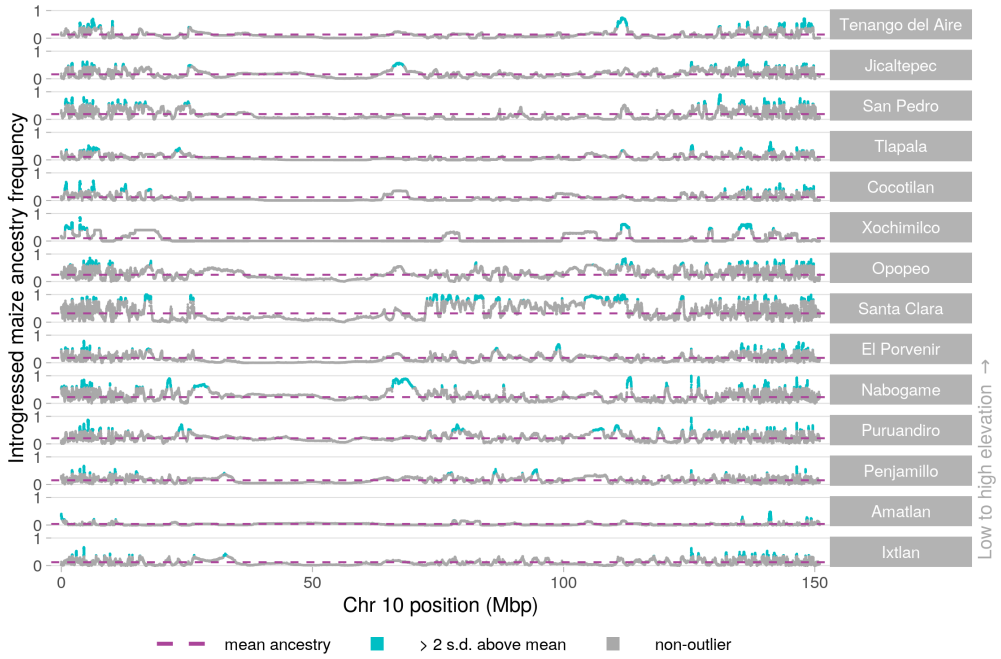

Fig S29. Introgression in *mexicana* populations across chromosome 10

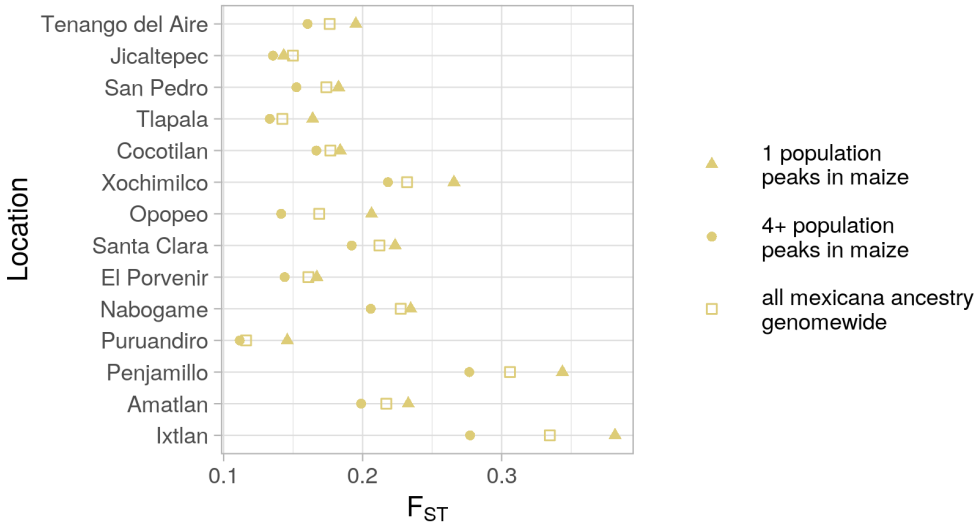

Fig S30. Differentiation (F_{ST}) between introgressed ancestry tracts and local *mexicana* Each point summarises F_{ST} between *mexicana* ancestry tracts within a focal maize population and *mexicana* ancestry tracts within the local *mexicana* population sampled at the same site. Within-*mexicana* 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. Notably, peaks where adaptive introgression is limited to the local population ('1 population peaks') do not have reduced F_{ST} to local *mexicana* haplotypes.

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.05.434040; this version posted September 21, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

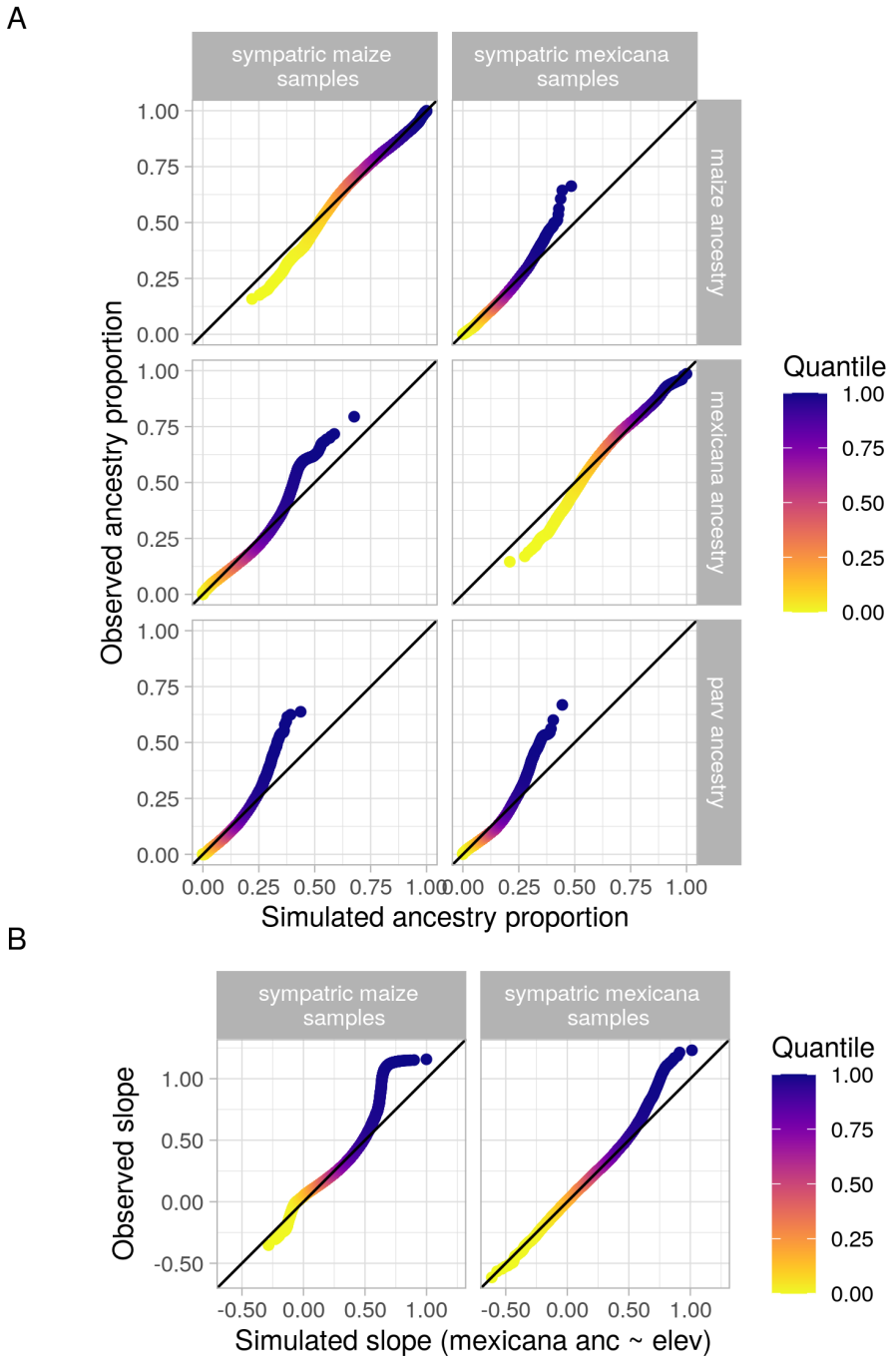

Fig S31. Quantile comparison of observed data vs. MVN normal null model (A) QQ-plot of simulated vs. observed mean ancestry at individual loci across all sympatric individuals. (B) QQ-plot of simulated vs. observed slopes from the linear model *mexicana* ancestry ∼ elevation at individual loci.

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.05.434040; this version posted September 21, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

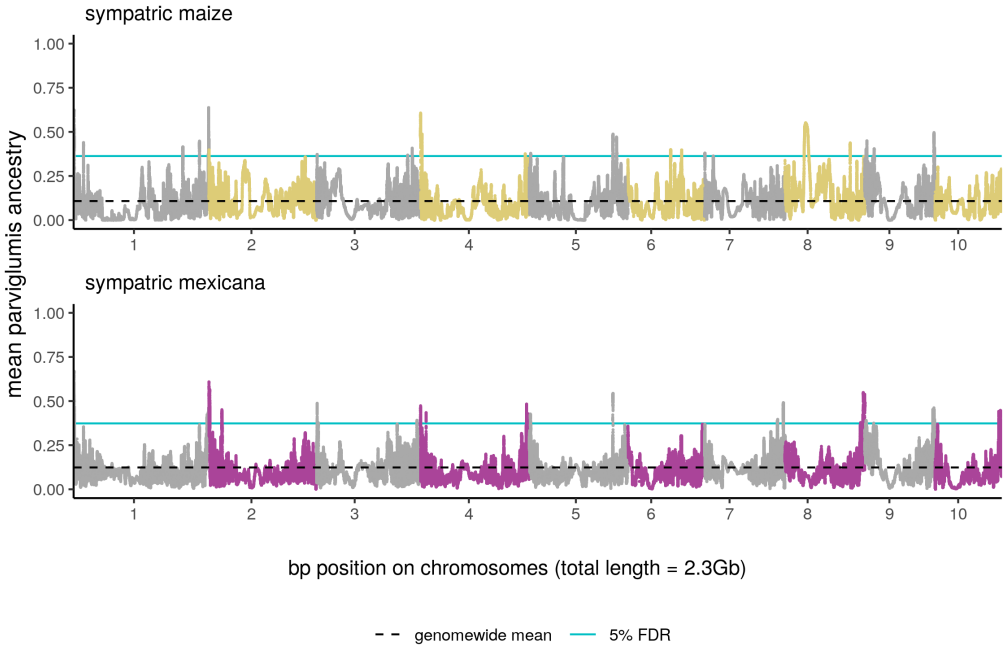

Fig S32. Genomewide scan for selection on *parviglumis* ancestry. Mean introgressed *parviglumis* ancestry into sympatric maize and *mexicana* populations populations. The blue line shows the 5% false discovery rate for high introgression, set using multi-variate normal simulations.

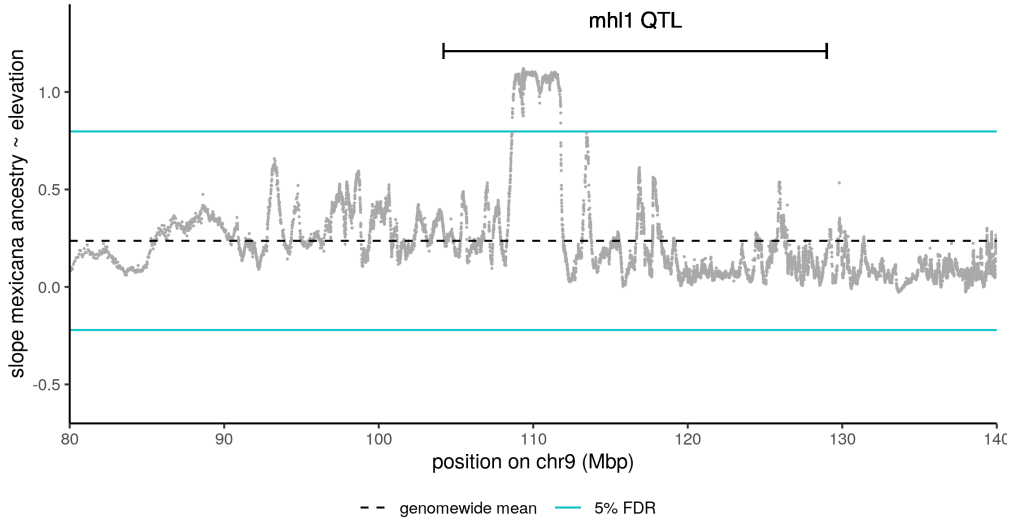

Fig S33. Ancestry slope with elevation at mhl1 locus. Slope of introgressed *mexicana* ancestry proportion in sympatric maize over a 1 km gain in elevation, zoomed in on the mhl1 QTL region on chromosome 9. Coordinates for the contiguous 3 Mb outlier region within this QTL are 9:108615836-111785557.

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.05.434040; this version posted September 21, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

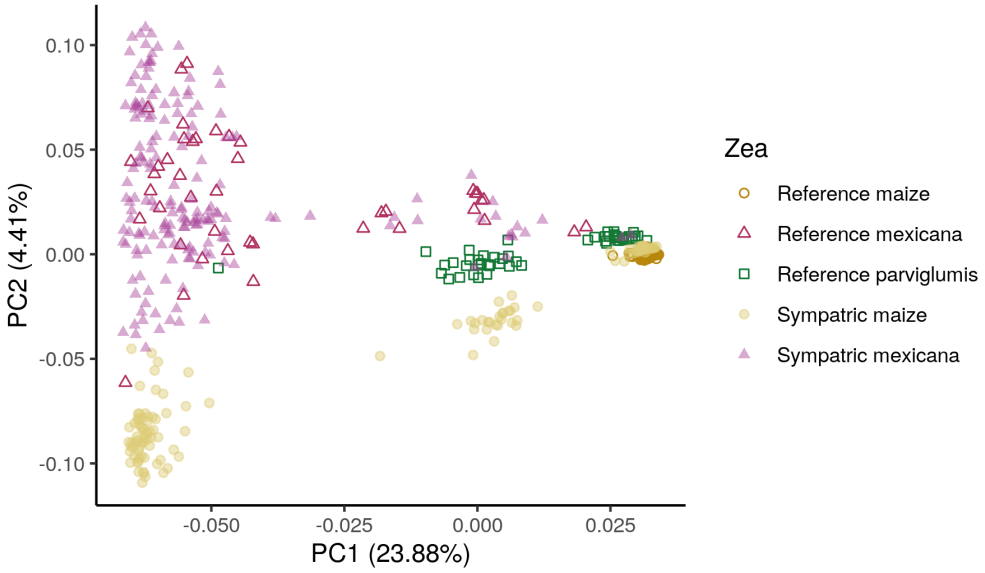

Fig S34. PCA of putative mhl1 inversion. Principal components analysis of all SNPs in the 3 Mb outlier region within the mhl1 QTL region that shows a steep increase in introgressed *mexicana* ancestry across elevation (<5% FDR). This region on chromosome 9 is a putative inversion (9:108615836-111785557), 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). There is evidence that the *mexicana* inversion allele segregates at low frequency in lowland *parviglumis*, but not lowland maize (right cluster includes all low-elevation reference maize). Additionally, PC2 primarily separates out diversity within the common *mexicana* allele for the inversion, with maize samples tending to have lower PC2 values.

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.05.434040; this version posted September 21, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

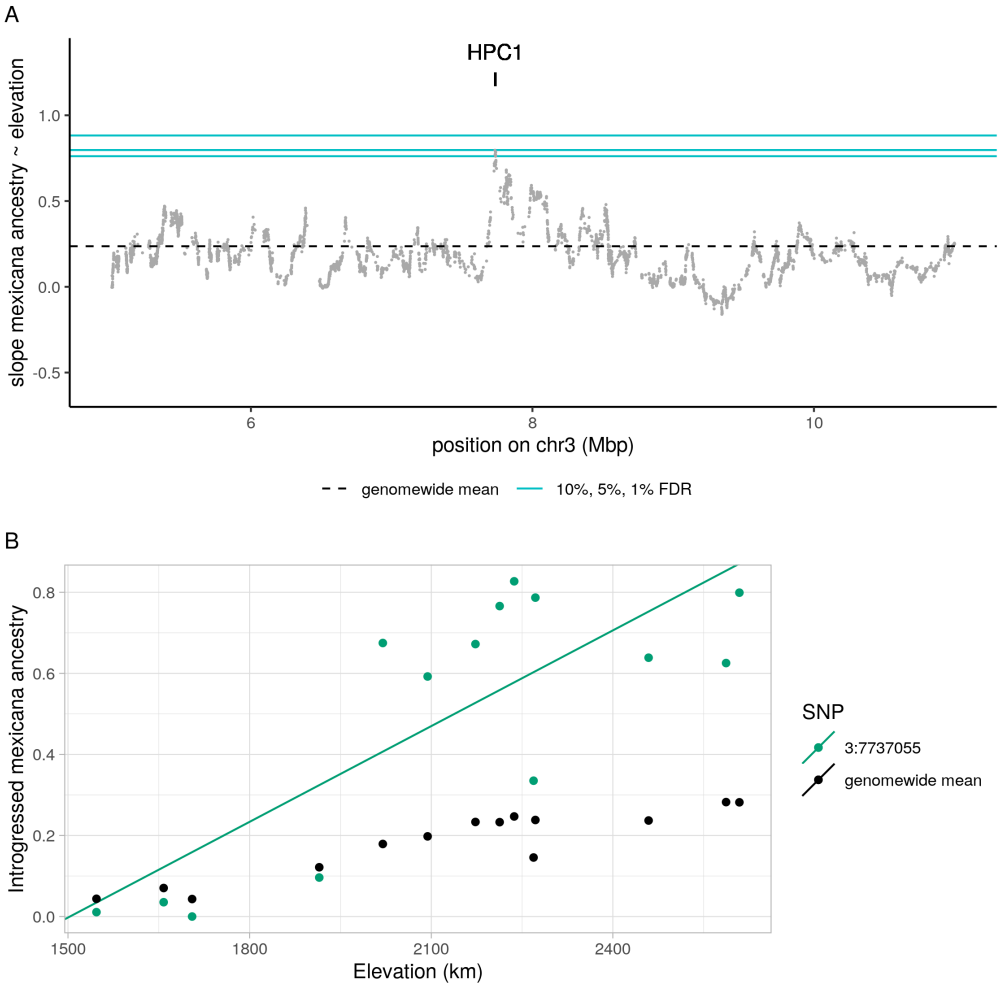

Fig S35. Introgression patterns at *HPC1* flowering time gene. (A) Slope of *mexicana* ancestry across elevation near the *HPC1* gene, with 1%, 5% and 10% FDRs for steep slopes indicated in blue. (B) Mean *mexicana* introgression proportions for each population, and fitted slope across elevation, for the top outlier SNP within *HPC1* (3:7737055). In higher elevation maize populations, introgression at *HPC1* far exceeds the genomewide mean *mexicana* ancestry (black circles).

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.05.434040; this version posted September 21, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

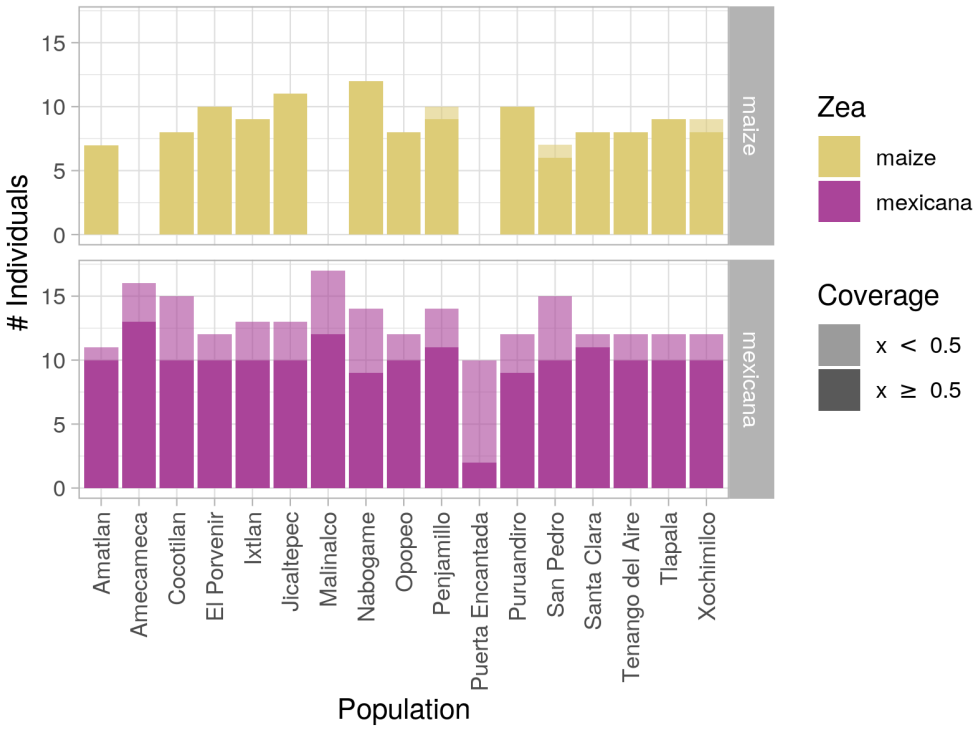

Fig S36. Number of individuals sequenced per location. Number of maize (top) and *mexicana* (bottom) individuals sequenced by this study with minimum 0.05x WGS coverage. Amecameca, Malinalco and Puerta Encantada have no paired maize samples and are used as a reference panel for *mexicana* ancestry. For 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.

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.05.434040; this version posted September 21, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

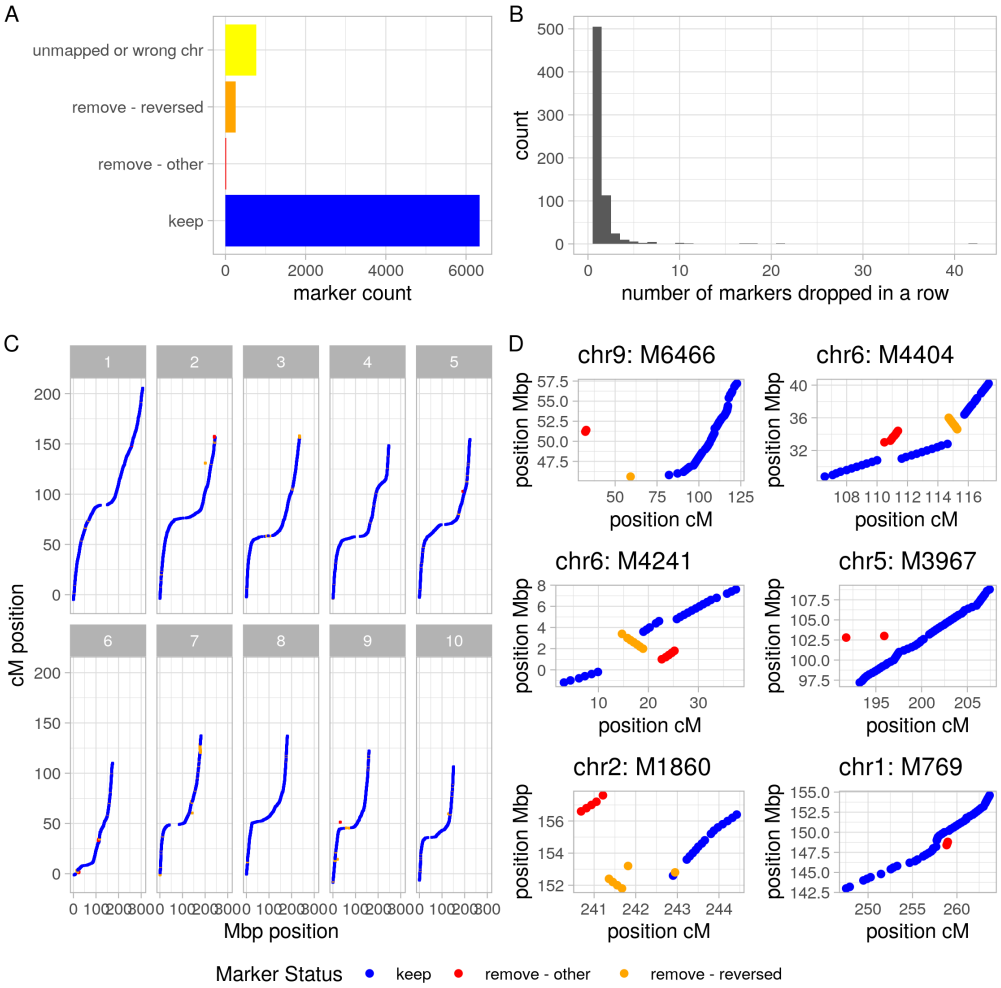

Fig S37. Linkage map. (A) Over 85% of the original markers from the Ogut et al [62] 0.2cM linkage map on reference genome v2 were successfully transferred to reference genome v4, using Assembly Converter (ensembl.gramene.org). We dropped markers automatically that mapped to the wrong chromosome or in reverse map order (presumably due to small contigs having corrected orientations in the newer version of the reference genome). A small number of markers were dropped by hand for more complex out-of-order mapping errors. (B) With few exceptions, the number of markers dropped in a row is small. (C) Visualization of the full linkage map on v4 of the reference genome. Included markers are in blue while excluded markers are highlighted in orange and red. (D) Zoomed in view of all 6 regions in the genome where out-of-order markers did not form simple reversals. For these more complex regions we identified markers to drop by hand (in red) to reach a monotonically increasing map solution.

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.05.434040; this version posted September 21, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

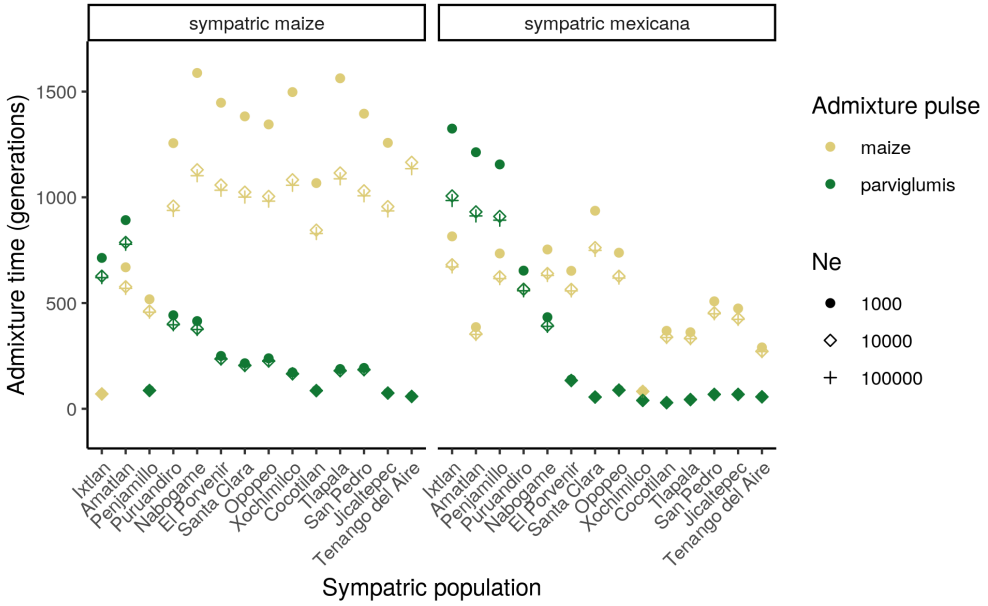

Fig S38. Sensitivity of timing estimates to choice of Ne. Admixture timing estimates from ancestry_hmm across 3 choices of effective population size. Timing estimates are strongly correlated with slightly older estimates for the smallest Ne = 1000. Local ancestry calls across the genome, summarised within sympatric maize and sympatric *mexicana* for each Ne, were all tightly correlated ($r > 0.99$). Main results of the paper are presented only for Ne = 10,000.