Molecular Parallelism Underlies Convergent Highland Adaptation of Maize Landraces

Li Wang,*,[1,2,3] Emily B. Josephs,[4,5] Kristin M. Lee,[3] Lucas M. Roberts,[2] Rubén Rellán-Álvarez,[6,7] Jeffrey Ross-Ibarra,*,[3,8] and Matthew B. Hufford*,[2]

[1]Shenzhen Branch, Guangdong Laboratory for Lingnan Modern Agriculture, Genome Analysis Laboratory of the Ministry of Agriculture, Agricultural Genomics Institute at Shenzhen, Chinese Academy of Agricultural Sciences, Shenzhen, China

[2]Department of Ecology, Evolution, and Organismal Biology, Iowa State University, Ames, IA, USA

[3]Department of Evolution and Ecology, University of California, Davis, Davis, CA, USA

[4]The Ecology, Evolution, and Behavior Program, Michigan State University, East Lansing, MI, USA

[5]Department of Plant Biology, Michigan State University, East Lansing, MI, USA

[6]Langebio, Irapuato, Gto., Mexico

[7]Department of Molecular and Structural Biochemistry, North Carolina State University, Raleigh, NC, USA

[8]Genome Center and Center for Population Biology, University of California, Davis, Davis, CA, USA

*Corresponding authors: E-mails: wangli03@caas.cn; rossibarra@ucdavis.edu; mhufford@iastate.edu.
Associate editor: Michael Purugganan

Abstract

Convergent phenotypic evolution provides some of the strongest evidence for adaptation. However, the extent to which recurrent phenotypic adaptation has arisen via parallelism at the molecular level remains unresolved, as does the evolutionary origin of alleles underlying such adaptation. Here, we investigate genetic mechanisms of convergent highland adaptation in maize landrace populations and evaluate the genetic sources of recurrently selected alleles. Population branch excess statistics reveal substantial evidence of parallel adaptation at the level of individual single-nucleotide polymorphism (SNPs), genes, and pathways in four independent highland maize populations. The majority of convergently selected SNPs originated via migration from a single population, most likely in the Mesoamerican highlands, while standing variation introduced by ancient gene flow was also a contributor. Polygenic adaptation analyses of quantitative traits reveal that alleles affecting flowering time are significantly associated with elevation, indicating the flowering time pathway was targeted by highland adaptation. In addition, repeatedly selected genes were significantly enriched in the flowering time pathway, indicating their significance in adapting to highland conditions. Overall, our study system represents a promising model to study convergent evolution in plants with potential applications to crop adaptation across environmental gradients.

Key words: convergent adaptation, flowering time, polygenic adaptation, population branch statistic.

Introduction

Convergent adaptation of populations to similar environments provides compelling evidence that natural selection, not neutral processes, shape trait variation. Similar phenotypes often arise independently in multiple species or populations exposed to the same evolutionary pressure (Davies et al. 2018; Alves et al. 2019). For example, species from four orders of insects, spanning 300 million years of divergence, have independently evolved tolerance to toxic compounds in milkweed and dogbane plant species (Dobler et al. 2012; Zhen et al. 2012). Likewise, reduction-of-function alleles at the *FLC* locus have evolved multiple times independently in *Capsella rubella* populations, conferring variation in flowering time (Yang et al. 2018). Pool et al. (2017) investigated cold adaptation across three pairs of highland and lowland *Drosophila*

melanogaster populations, finding strong evidence for alleles that were repeatedly selected during highland colonization.

But while convergent phenotypes have been observed across the tree of life, in many cases their underlying genetic basis (e.g., molecular parallelism in which the same nucleotides, genes, or pathways are targeted versus convergence through independent molecular means) is unknown (Rosenblum et al. 2014; Storz 2016). Study of the genetics of convergence can help shed light on fundamental questions in evolutionary biology, including whether natural selection is constrained and repeatable or instead characterized by many molecular paths to similar phenotypes. The answer to these questions may depend, to a certain extent, on the trait itself. For example, the genetic architecture of a phenotype is a primary determinant of whether trait convergence results

Open Access

Article

from parallel selection on orthologous loci (Yeaman et al. 2018). Such convergence is more likely for simple traits where only a few loci contribute or when loci are subject to antagonistic pleiotropy, and less likely for highly polygenic traits such as biomass or height (Orr 2005; Rosenblum et al. 2014; Yeaman et al. 2018).

Convergence through molecular parallelism, when it occurs, can originate in a number of ways. Genetic variation underlying convergent traits can arise as independent mutations, be derived from existing standing variation in a shared ancestral population, or be transferred between populations via migration (Lee and Coop 2017). Empirical studies have documented each of these modes of parallel adaptation. For instance, independent evolution of C4 photosynthesis in grasses (Christin et al. 2007) and sedges (Besnard et al. 2009) has involved multiple de novo mutations in the phosphoenolpyruvate carboxylase gene, a key C4 enzyme. In contrast, parallel adaptation in rabbit populations resistant to the myxoma virus in Australia, France, and the United Kingdom is achieved through selection on standing variation in immunity-related genes (Alves et al. 2019). Finally, convergence in patterns of wing coloration in butterfly species, an example of Müllerian mimicry, has resulted from gene flow across species (Dasmahapatra et al. 2012; Edelman et al. 2019).

Given its history, maize is an ideal model system to study parallel adaptation. Maize was first domesticated in the warm lowlands of the Balsas River Valley approximately 9,000 years ago (Matsuoka et al. 2002; Piperno et al. 2007), and subsequently spread to several independent highland regions: the Mexican Central Plateau, the highlands of the southwestern United States and Guatemala, and the Andes (Piperno and Flannery 2001; Dickau et al. 2007; Merrill et al. 2009; Piperno et al. 2009; Grobman et al. 2012; Hufford et al. 2012; da Fonseca et al. 2015; Bush et al. 2016; Stephen Athens et al. 2016; Brandenburg et al. 2017; Emerson et al. 2020; Wang et al. 2020). Although highland regions colonized by maize are far from identical, commonalities include a shorter growing season, low temperature, low partial pressure of atmospheric gases, and high ultraviolet radiation (Korner 2007; Lomax et al. 2012). Highland individuals exhibit several traits that are thought to be adaptive under these conditions, including highly pigmented and hairy leaf sheaths (Lauter et al. 2004; Rodriguez-Zapata et al. 2021). Common garden experiments have also demonstrated that highland maize flowers substantially earlier than lowland material (Eagles and Lothrop 1994; Jiang et al. 1999; Rodriguez et al. 2006).

The genetic basis of highland phenotypes across these populations is largely unknown, however. Earlier population genetic analysis of maize in Mexico and the Andes identified a small percent of single-nucleotide polymorphism (SNPs) showing evidence of selection in both populations, but concluded that maize highland adaptation arose largely independently in the two populations (Takuno et al. 2015). One contributor to the differences between these populations has been gene flow from wild relatives: adaptive introgression from the wild congener teosinte (Zea mays ssp. mexicana, hereafter mexicana) is thought to have played an important role during highland adaptation in Mexico (Hufford et al. 2013; Calfee et al. 2021; Rodriguez-Zapata et al. 2021), but maize has no wild relatives in South America and simulations suggested that long-distance gene flow between the populations is unlikely (Takuno et al. 2015). Nonetheless, this earlier work was limited to two highland populations and relied on limited genotyping data. Subsequent investigation using whole genome sequencing, for example, has discovered alleles introgressed from mexicana in both the Guatemalan and southwestern U.S. highlands outside of the distribution of mexicana (Wang et al. 2017), suggesting that migration from the Mexican highlands may not be as implausible as previously thought.

Here, we evaluate evidence for trait convergence through parallel molecular means using resequencing data from four highland populations of maize (Southwestern United States, Mexico, Guatemala, and the Andes) and two lowland populations (Mexico and South America). We assess the prevalence of molecular parallelism at the SNP, gene, and pathway levels, comparing models of de novo mutation, standing variation, and gene flow. Finally, our analysis of parallelism in pathways focuses on the well-characterized flowering time genes in maize and assesses signatures of polygenic adaptation for this trait.

Results

Description of Samples and Data

Four independent highland populations (the Southwestern United States [hereafter referred to as US, six samples], Mexico [MH, six samples], Guatemala [GH, three samples], and the Andes [AN, five samples]) and two reference lowland populations (Mexico [ML, five samples] and South America [SL, six samples]) were sampled to investigate highland adaptation in maize landraces (fig. 1 and supplementary table S1, Supplementary Material online). SNPs called from high-depth, whole-genome resequencing data were obtained from our previous study (Wang et al. 2017). After filtering (see Materials and Methods), a total of 1,567,351 SNPs across 35 samples were retained for further analyses. We also characterized how environmental factors varied across our highland and lowland populations. A principal component analysis (PCA) of 19 bioclim environmental variables (http://www.worldclim.org/) revealed that highland and lowland accessions were differentiated along PC1 (comprising 51.8% of variation); whereas PC2 (comprising 22.8% of variation) reflected latitudinal differences across highland populations (supplementary fig. S2, Supplementary Material online). By plotting the loadings from the 19 bioclim variables, we found PC1 was primarily polarized by temperature seasonality and PC2 by precipitation (supplementary fig. S2, Supplementary Material online).

Convergence through Parallelism at the SNP Level

In order to detect loci targeted by selection in highland populations, we utilized the Population Branch Excess (PBE) statistic (Yassin et al. 2016), which characterizes changes in allele frequencies in a focal population relative to two independent "outgroup" populations (Pool et al. 2017). PBE values were
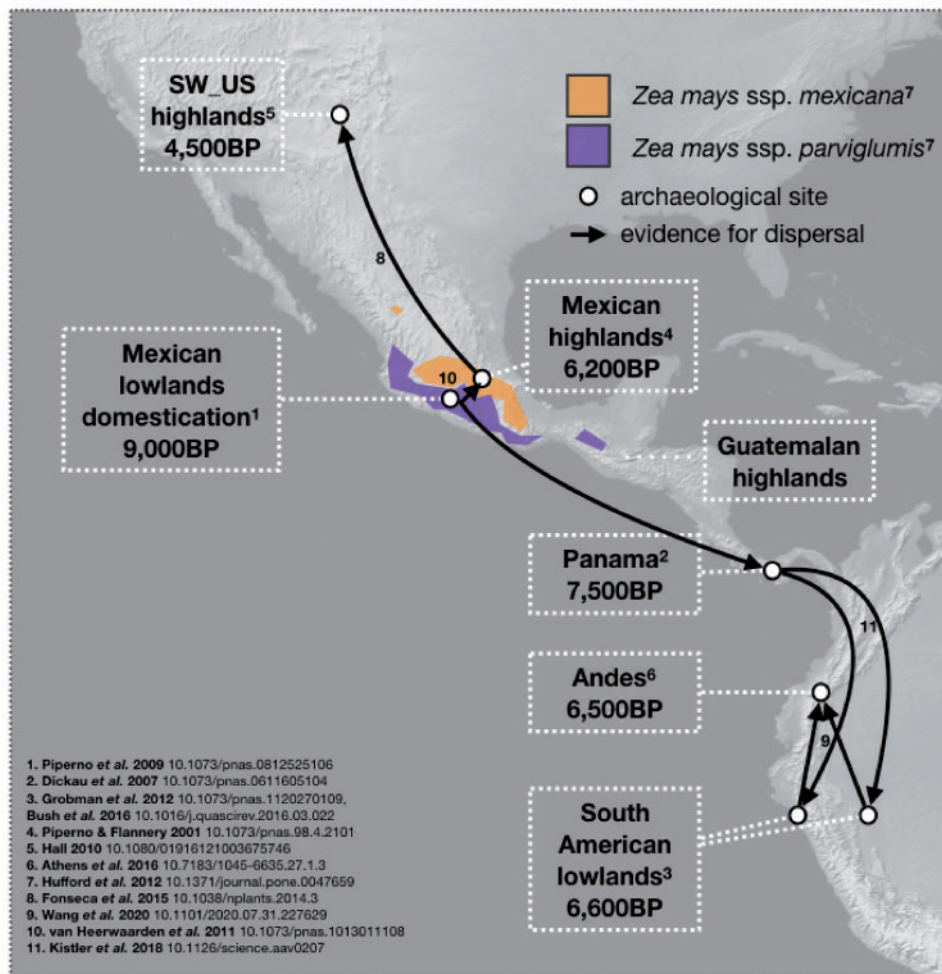
FIG. 1. Sampling locations and expansion route of maize landraces, modified from Ross-Ibarra and Piperno (2020). Domestication and expansion times for maize populations are from published articles (Piperno and Flannery 2001; Dickau et al. 2007; Piperno et al. 2009; Hall 2010; van Heerwaarden et al. 2011; Grobman et al. 2012; Hufford et al. 2012; da Fonseca et al. 2015; Bush et al. 2016; Stephen Athens et al. 2016; Kistler et al. 2018; Wang et al. 2020). The exact location of our samples is indicated in supplementary fig. S1, Supplementary Material online.

calculated for each SNP site and the top 5% were considered outliers and potential targets of selection.

To gauge the extent of convergence across our highland populations at the SNP level, we evaluated whether shared *PBE* outliers in pairs of populations and larger groups were more common than expected by chance. Significant overlap was observed in all pairs of highland populations ($p \approx 0$, hypergeometric statistic test; supplementary tables S2 and S3, Supplementary Material online). The Mexican Highland population showed the most substantial overlap in outlier SNPs with the Guatemalan Highlands (13,461 overlapped SNPs among 53,944 putatively selected SNPs in the Mexican Highlands; 7.5-fold enrichment) and the Southwestern U.S. Highlands (10,884 overlapped SNPs among 53,944 putatively selected SNPs in the Mexican Highlands; 6-fold enrichment). Although significantly more than expected, shared outliers between the Andes and all other highland populations were not nearly as extensive (4,031–6,702 overlapped SNPs among 51,453 putatively selected SNPs in the Andes; between 2- and 4-fold enrichment), consistent with reduced parallelism at the SNP level between these isolated

regions (supplementary table S2, Supplementary Material online). The intersection of outlier SNPs across all highland populations was also significantly enriched (supplementary fig. S3, Supplementary Material online). Enrichment of overlapping selected sites was also confirmed in a more stringent tail of the top 1% of *PBE* values (supplementary fig. S4, Supplementary Material online) and in a data set that was thinned to account for linkage disequilibrium (supplementary fig. S5, Supplementary Material online). Although we did observe significant enrichment for parallelism at the SNP level, it is important to note that the majority of putatively selected SNPs (from 59.7% to 68.1%) showed signatures of adaptation in a single highland population.

Our analysis of convergence at the SNP level using *PBE* revealed some particularly compelling candidate loci. For example, a SNP within *PIF3.1* (phytochrome-interacting factor) exhibited the highest *PBE* value in the Mexican Highland population and was also detected as a target of selection in all other highland populations (fig. 2A). A nonsynonymous, derived allele at this locus is fixed across all highland populations (fig. 2C). In addition, SNPs within *GRMZM2G078118*, a
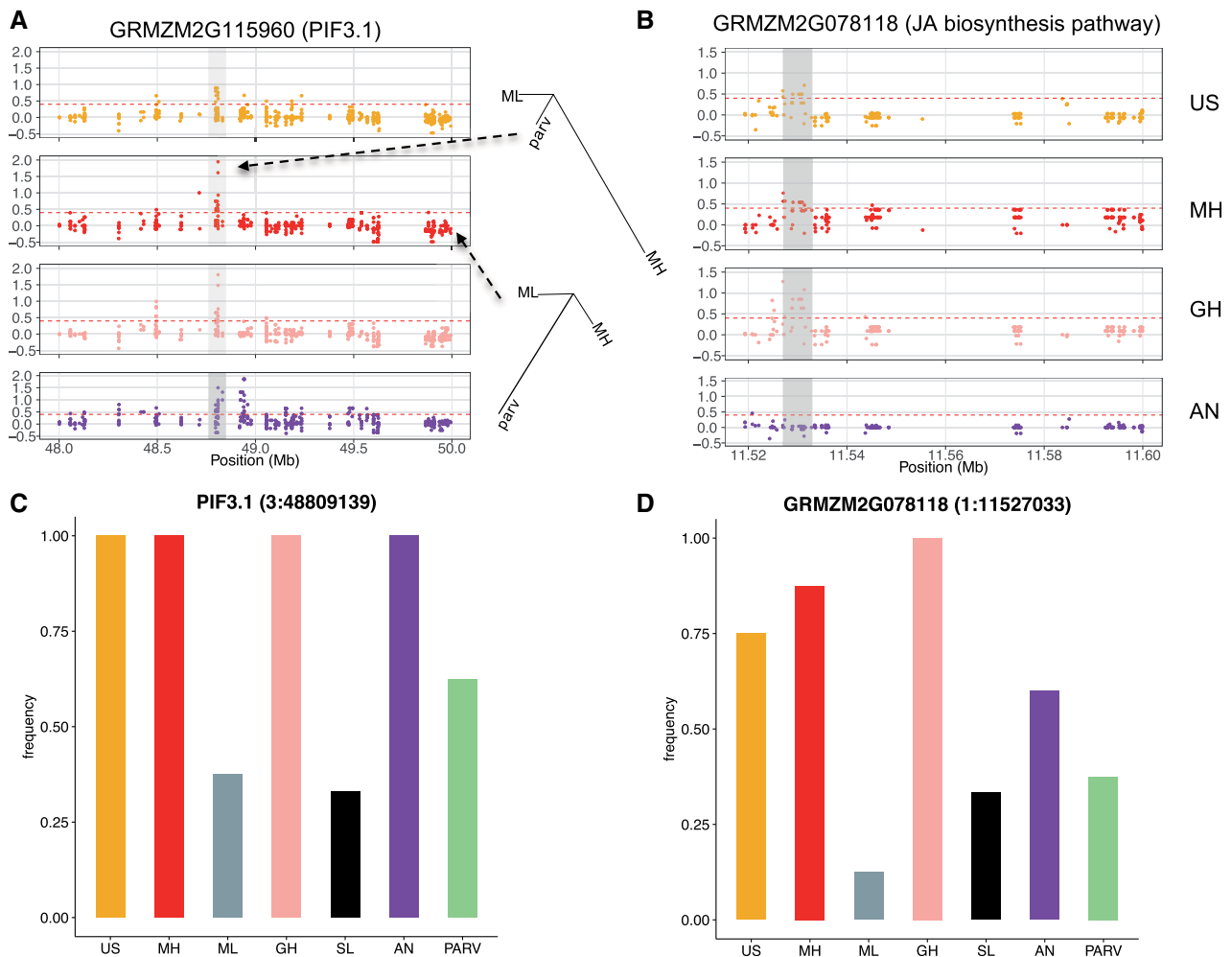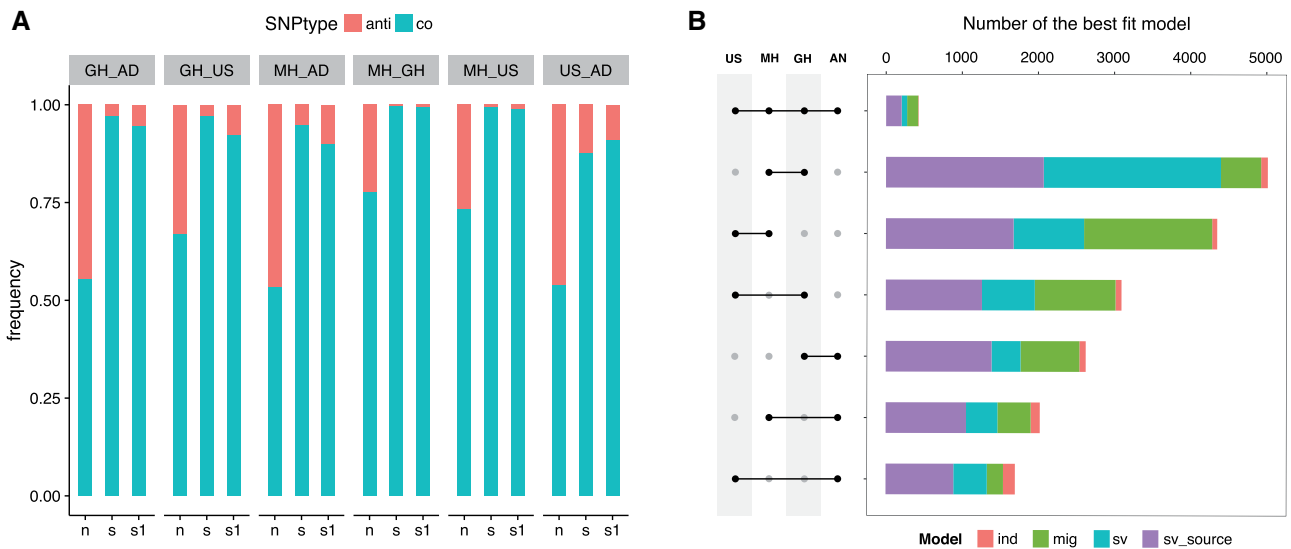
FIG. 2. Patterns of parallel adaptation at two candidate loci. (*A*) Distribution of *PBE* values in a 2-Mb region around the gene *PIF3.1*. The branch length of inset trees was based on *PBE* values and indicates the difference between a selected and unselected SNP. (*B*) Distribution of *PBE* values in a 0.5-Mb region around the gene *GRMZM2G078118* involved in jasmonic acid biosynthesis. (*C*) Barplot of the reference allele frequency of one SNP located in *PIF3.1*. (*D*) Barplot of the nonreference allele frequency of one SNP located in *GRMZM2G078118*.

gene included in the jasmonic acid biosynthesis pathway, were detected as outliers in all Mesoamerican highland populations (Southwestern US, Mexican Central Plateau, Guatemalan ighlands), but not the Andes (fig. 2B and D).

Although our *PBE* analysis reveals that selection targeted many of the same SNPs for adaptation across highland populations, this test alone does not clarify whether the same specific alleles were adaptive. We term outlier SNPs with the same allele elevated to high frequency in highland population pairs "codirectional," and SNPs with different alleles as "antidirectional." Predominantly, shared outliers in highland population pairs show codirectional allele frequency shifts (> 95% in all population pairs with the exception of the Southwestern US/Andes comparison, which was 87.6%). These results provide stronger evidence of molecular parallelism at the SNP level (fig. 3A). In contrast, random samples of SNPs that do not show evidence of selection exhibit much reduced signals of codirectional change. Although random SNP samples between the Andes and other highland populations show roughly equal proportions of co- and antidirectionality, substantially more codirectional SNPs are observed in

comparisons of Mesoamerican populations, perhaps reflecting their more closely shared histories (fig. 3A). However, the strong excess of codirectional changes in candidate SNPs may also be influenced by ancestral frequencies in lowland populations. For example, neutral drift has a greater likelihood of simultaneously fixing the same allele in multiple highland populations if the allele is already at high frequency in the lowlands. In order to address this bias, we first approximated the ancestral allele frequencies of outlier SNPs (based on the two-dimensional site frequency spectrum of Mexican lowland maize and the wild relative *Z. mays* ssp. *parviglumis*) in a subset of the neutral SNP set, and recalculated the number of codirectional and antidirectional SNPs. The same pattern was revealed (supplementary fig. S6, Supplementary Material online). Second, we subsampled outlier SNPs with a high minor allele frequency (i.e., between 0.3 and 0.5) and recalculated the ratio of codirectional versus antidirectional SNPs. We found antidirectional SNPs were slightly increased in this subset, but codirectional SNPs were still far more numerous (fig. 3A). Taken together, these results suggest that the observed codirectionality of allele frequency change in shared outlier SNPs provides strong evidence

FIG. 3. Patterns of parallel adaptation across loci. (A) Distribution of codirectional and antidirectional SNPs in neutral and outlier SNP sets. n: neutral SNPs; s: common outliers; s1: a subset of common outliers with MAF between 0.3 and 0.5. (B) Number of best fit models for genetic source of repeatedly selected SNPs. ind: independent de novo mutation; mig: migration; sv: standing variation; sv_source: standing variant with a source population. Abbreviations for populations: AN, Andes; GH, Guatemalan highlands; MH, Mexican highlands; US, Southwestern US highlands.

of molecular parallelism at the SNP level among highland maize populations.

The Source of Shared Adaptation among Highland Populations

To determine whether shared *PBE* outlier SNPs had independent origins in highland populations, we applied DMC, a composite-likelihood-based inference method that distinguishes among possible mechanisms (Lee and Coop 2017; Tittes 2020). We controlled for possible effects of linkage disequilibrium by subsampling outliers to one SNP per 2-kb window. First, we examined SNPs selected in all four independent highland populations. We found the vast majority of these common outliers (421 among 428) had the highest composite likelihood under models where there was a single origin of the beneficial allele. Within this subset, 47.4%, 34.1%, and 16.8% of outliers had the standing variant model from a source population, the migration model and the standing variation model as best fits, respectively (fig. 3B). The standing variant model from a source population (outlined in detail in Appendix A.4 of Lee and Coop 2017) specifies that a beneficial allele originates in a single source population then spreads via ancient gene flow to other adapted populations where it may segregate for a standing time t prior to the onset of selection. Among shared outliers assigned the standing variant model from a source population, 70.9% and 25.6% had the second-most-likely model as the standing variation and migration models, respectively, with 82.3% of these showing likelihood value differences <1 between the best-fit and second-best-fit models. Considering the potential for overfitting with the standing variant model with a source population, which has four parameters compared with three parameters in the migration and standing variant models, we propose that the majority of common outliers assigned as the standing variant with a source model were virtually indistinguishable from the standing variant and migration models. For the common outliers identified under the recent migration model, the main migration sources were the Mexican (43.9%) and Guatemalan (41.9%) highland populations, a result that could be partially explained by the substantial shared ancestry of these two populations and direct maize colonization from the Mexican highlands to the highlands of both the Southwestern US and Guatemala. In summary, results for outlier SNPs shared across all four highland regions suggest gene flow among highland populations and standing variation may have been the major sources of adaptive variation.

We further explored the source of shared adaptive variation across pairwise samples of highland populations. As seen in four-way comparisons after correction, the standing variant with a source population model (38.5–52.9%; mean 46.5%), the standing variation model (14.5–46.4%; mean 24.0%), and the migration model (10.5–38.7%; mean 25.9%) were most commonly supported (fig. 3B). The standing variant with a source model, which includes an extra parameter, showed only a slightly higher likelihood than the simple standing variation and migration models, indicating these sources may also be likely drivers of convergence. It is noteworthy that the parallelism level among Mesoamerican highland populations was much stronger than observed between the Andes and other highland populations (fig. 3B). Not only do fewer loci show evidence of selection in pairs of populations including the Andes, but repeatedly selected loci are more likely to have arisen independently: 6.1% of shared loci were identified as having independent origins in Andean comparisons, in contrast to a mean of 1.9% within Mesoamerican highland population comparisons (fig. 3B).

Previous studies have suggested that migration between the Andes and Mesoamerican populations is unlikely (Takuno et al. 2015). We therefore further evaluated all

repeatedly selected SNPs by calculating the divergence statistic, F_{ST}, between pairwise highland populations to assess how this varies based on the assigned model. Repeatedly selected SNPs consistent with a migration model exhibited weaker divergence (lower F_{ST}; $p \approx 0$; supplementary fig. S7, Supplementary Material online), an additional indicator of gene flow. We also calculated the $\widehat{f_{dM}}$ statistic (Malinsky et al. 2021) in 10-kb nonoverlapping windows across the genome. In a $(((P1, P2)P3),O)$ shaped phylogenetic tree, the $\widehat{f_{dM}}$ statistic gives positive values for introgression between P3 and P2 and negative values for introgression between P3 and P1. We found SNPs consistent with the migration model demonstrated more positive $\widehat{f_{dM}}$ values, providing additional support for gene flow between the Andes and Mesoamerican highland populations (supplementary fig. S8, Supplementary Material online). In summary, although the proportion of dually selected SNPs between the Andes and Mesoamerican highland populations (from 7.8% to 11.2%) was consistent with previous work (7–8%) (Takuno et al. 2015), the two most supported source models were standing variation and migration, suggesting a more prominent role for migration than previously hypothesized (Takuno et al. 2015).

Molecular Parallelism at the Genic Level

We next evaluated the extent of molecular parallelism at the gene level. A gene was classified as an outlier if it contained PBE outlier SNPs within the gene or 10 kb upstream or downstream (supplementary table S4, Supplementary Material online). A total of 1,651 candidate genes, among the total of 29,160 genes with sufficient sequence depth to be tested, were observed in the Mexican Highlands, among which 360 (21.8%) contained or were near to outlier SNPs which were shared with at least one other highland population. The percentage dropped to 18.9% in the Guatemalan highlands, 10.0% in the Southwestern U.S. highlands and 9.4% in the Andes (supplementary table S5, Supplementary Material online). We utilized the hypergeometric test implemented in the R package SuperExactTest (Wang et al. 2018) to evaluate the two- to four-degree intersection of outlier genes among highland populations. As observed in results at the SNP level, significant overlap was observed for each comparison (supplementary figs. S9 and S10, Supplementary Material online). In order to account for the effects of linkage disequilibrium on selected genes, we thinned outlier SNPs with $r^2 \leq 0.2$ within 10-kb windows and reevaluated the overlap of selected genes. This thinned set showed consistent results (Supplementary Material online), suggesting the pattern was not caused by clustered outlier genes in selective sweeps. In many instances, common outlier genes involved selection on different SNPs across highland populations. For example, 349 (21.1%) of the 1651 selected genes in the Mexican highlands were in common with other highland populations, but showed evidence of selection on distinct SNPs. The percentage was 18.8% in Guatemala, 24.7% in the Southwestern US, and 18.5% in the Andes (supplementary table S5, Supplementary Material online).

Constraint of the Adaptation Target Size Leads to Molecular Parallelism

There are two explanations for the extent of molecular parallelism we have observed in genes selected in the highlands (Yeaman et al. 2018). First, it is possible that adaptation is constrained by the number of genes that can affect a trait, effectively placing physiological limits on the routes adaptation can take. Alternatively, many different genes may have the potential to affect the trait, but deleterious pleiotropic effects of variation in some genes may prevent them from playing a role in adaptation. We will refer to these possibilities as "physiological" and "pleiotropic" constraint. Consistent with our analysis above, Yeaman et al. (2018)'s C_{χ^2} statistic finds strong evidence that selection at the same loci is shared among populations more often than can be explained by chance (C_{χ^2} from 19.6 to 40.1 with a mean of 28.05, permutation $P < 0.0001$). To further evaluate constraint in parallel adaptation, we tested an alternative null model which assumes the set of genes selected in at least one population represents all genes that can possibly affect the trait. We then tested whether there was evidence of excessive sharing beyond the expectations under physiological constraint (Yeaman et al. 2018). Here, we found that C_{χ^2} was less than 0 in all highland pairs except the Mexican and Guatemalan highlands pair ($C_{\chi^2} = 1.9$, $P = 0.025$), indicating sharing beyond that expected under physiological constraint and consistent with the known substantial shared ancestry between these two highland populations. These results indicate that repeated selection of genes among most highland pairs may simply be due to physiological constraint in the number of loci that can contribute to variation in a trait. To explore this possibility further, we assessed the prevalence of molecular parallelism in the flowering time pathway, a well-characterized gene network known to be of adaptive importance in highland environments.

Molecular Parallelism within the Flowering Time Pathway

Flowering time in maize is known to be a highly quantitative trait and an example of polygenic adaptation (Buckler et al. 2009; Navarro et al. 2017; Gates et al. 2019; Josephs et al. 2019). As such, we may expect that highland adaptation for flowering is the result of subtle, coordinated allele frequency shifts across many loci. In evaluating flowering time genes we tested: 1) the strength of selection on loci associated with flowering time relative to expectations based on drift, and 2) the intersection of selected genes in the pathway in multiple highland regions. To first identify whether flowering time was targeted during highland adaptation, we conducted a genome-wide association study (GWAS) of 29 traits measured in short-day conditions in Ponce, Puerto Rico and Homestead, Florida (Hung et al. 2012) using a maize association mapping panel comprised of tropical and temperate inbred lines (Flint-Garcia et al. 2005). We then summarized the frequency of multiple alleles affecting a trait of interest with a polygenic score and tested for associations between polygenic score and elevation of origin of our landraces, while

controlling for variation in relatedness between samples following Josephs et al. (2019). A significant result implies that allele frequencies that increase trait value are more common at one end of the elevation gradient (either high or low) than expected due to drift. Seven traits showed a significant relationship between their conditional polygenic score and elevation with a false discovery rate < 0.1. These traits were primarily related to flowering time: Days to Silk, Growing Degree Days to Silk, and Ear Height from '06 Puerto Rico, Growing Degree Days to Silk, Growing Degree Days to Tassel, and Ear Height in '06 Florida, and Growing Degree Days to Silk in '07 Florida (fig. 4).

Given this support for selection on flowering time during highland adaptation, we explored the overlap between our selected genes and known flowering time candidates. First, we utilized a list of 904 maize flowering time candidate genes aggregated by Li et al. (2016) (supplementary table S6, Supplementary Material online). Within this list, we found a substantial excess of flowering time genes targeted by selection among groups of two, three, and four highland populations ($p < 2.87e - 5$, hypergeometric test; fig. 5A and supplementary table S7, Supplementary Material online). Outlier SNPs within these genes also showed strong codirectional changes in allele frequency across highland populations ($91.8 - 100\%$) (supplementary fig. S11, Supplementary Material online). As observed in our genome-wide analysis, candidate sharing at flowering time loci was stronger among Mesoamerican highland populations than between comparisons with the Andes (fig. 5). We also applied the C_{χ^2} statistic to this list of genes and confirmed enriched sharing of selected genes among highland populations (mean C_{χ^2} 6.618, permutation $P < 0.0001$). Our finding of an excess of shared candidates across regions within a large set of known genes that affect flowering time suggests physiological constraint alone cannot explain patterns of convergence, but instead that variation at only a subset of genes can affect flowering without deleterious pleiotropic consequences.

We then narrowed our focus to a subset of core flowering time genes (Dong et al. 2012) with highly characterized functional roles (fig. 5C) based on published QTL or GWAS analyses. Thirty-two of the 39 core flowering time genes (contained at least one outlier SNP within the gene or in the 10-kb flanking regions) in at least one highland population, which indicated overrepresentation of flowering time genes among all selected genes ($P = 0.005$, hypergeometric test). Although common outlier genes between the Andes and other highland populations were not significantly enriched in this smaller list ($p > 0.074$, hypergeometric test; supplementary table S7, Supplementary Material online), enrichment was observed among all other two and three highland population comparisons ($P < 0.042$, hypergeometric test; supplementary table S4, Supplementary Material online). These results were confirmed through calculation of the C_{χ^2} statistic (Mesoamerican highland population comparisons, mean C_{χ^2} 2.32, permutation $P < 0.01$; comparisons to the Andes, mean C_{χ^2} 1.575, permutation P from 0.015 to 0.054).

The PBE values of outlier SNPs within or in the 10-kb flanking regions of flowering time genes were in the extreme tail of the distribution of all genes (fig. 5B), with some of the best-characterized flowering time genes in maize clearly targeted by selection in multiple highland populations (fig. 5C). For example, ZCN8 (GRMZM2G179264), which is orthologous to FLOWERING LOCUS T (FT) in Arabidopsis, has been shown to play a significant role in maize adaptation to high latitudes (Guo et al. 2018) and was selected in all four highland populations in our study. Floral transition is regulated by several MADS-box genes, particularly ZMM4 (GRMZM2G032339), which is under selection in all but the Southwestern U.S. highlands. VGT1 (GRMZM2G700665; vegetative to generative transition 1) was identified as an outlier in all four highland populations, with the same nonsynonymous SNP (8:131578990, indicating locus 131578990 on chromosome 8) targeted in the Mexican and Guatemalan highlands (supplementary fig. S12, Supplementary Material online). In contrast, in the Southwestern U.S. highlands, the selected SNP (8:131580179) in VGT1 was located in the 3′ UTR region and, in the Andes, the selected SNP (8:131579463) was a synonymous mutation near the 3′ UTR (supplementary fig. S12A–C, Supplementary Material online). Two CONSTANS (CO) genes, CONZ1 and ZmCCT, were selected in both the Mexican and Guatemalan highland populations. We further found Gigantea2 (GRMZM5G844173) was selected in all but the Guatemalan highlands and FKF2 (GRMZM2G106363) in all but the Mexican highlands. Among circadian clock genes, LUX (GRMZM2G067702) (supplementary fig. S12D, Supplementary Material online) was selected in all four highland populations and ZmPRR59 was an outlier in the three Mesoamerican highland populations. In addition, the F box protein ZEITLUPE (GRMZM2G113244), RVE2 (GRMZM2G145041) and PRR5 (GRMZM2G179024) were selected in all four highland populations. Finally, the light transduction gene PHYA1 (GRMZM2G157727) was targeted by selection in all but the Southwestern U.S. highland populations.

Discussion

The prevalence of parallel evolution is a long-standing question in the field of evolution. Through a comprehensive evaluation of parallel adaptation among independent maize highland populations, we found that, whereas most adaptation is independent, parallelism is more common than expected. We observed a particularly high level of parallelism among Mesoamerican highland populations relative to comparisons including Andean maize. The vast majority of parallel adaptive alleles we discovered have risen in frequency in the highlands from migration and standing variation, and only a small percentage represent de novo mutations. However, the proportion of repeatedly selected SNPs from de novo mutations is the highest in pairs of populations that include the Andes. Our analyses further reveal that adaptive routes from genotype to phenotype have been canalized by both physiological (i.e., few genes contributing to a trait) and pleiotropic (i.e., a subset of causal genes can contribute to trait variation without negative pleiotropic consequences) processes. For example, we observe that only a subset of possible genes
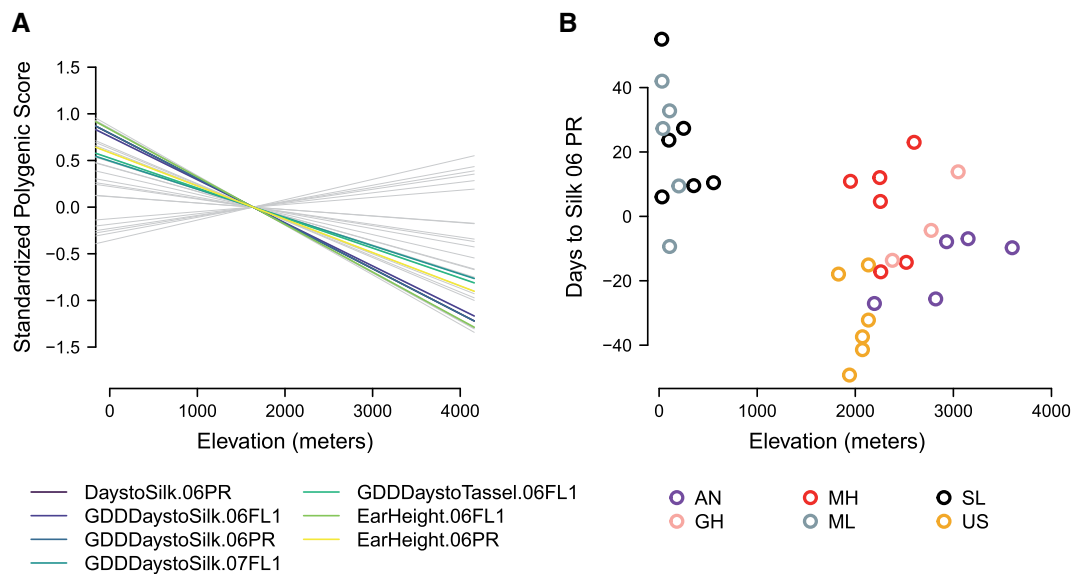
Fig. 4. Polygenic adaptation along elevation in the landraces. (A) A linear regression between polygenic score for all 29 trait-environment combinations tested and elevation of origin. The lines for the seven traits that showed significant signals of polygenic adaptation are colored and all other traits are shown in gray. (B) Polygenic score for days to silk in Florida (2006) for all landraces is negatively correlated with elevation of origin.

related to flowering time contribute to highland adaptation, with many of these loci showing evidence of selection in multiple highland regions, particularly in Mesoamerica.

Approximately one quarter (mean 24.0%) of highland adaptation SNPs showed a pattern of molecular parallelism in Mesoamerica. A smaller percentage also showed this pattern in Andean comparisons (mean 10.7%), but this represents a slightly higher proportion than previously reported (7–8% between the Andes and the Mexican Highlands) (Takuno et al. 2015). We have also expanded upon previous studies of parallel highland maize adaptation (Takuno et al. 2015) in a number of ways. Our study has broadened the geographic scope of sampling for highland maize populations and is the first such study using high-depth, whole-genome resequencing data, which allowed us to attain a more complete picture of parallelism. Additionally, our use of the *PBE* statistic allowed for detection of selection specifically in the highlands and reduced the extent of false positives, a limitation of studies that focus solely on population differentiation outliers. Our polygenic adaptation analyses also helped to clarify specific phenotypes targeted by selection during adaptation, showing clear evidence of selection on flowering time. We have also uncovered a pervasive role of migration in transferring adaptive alleles across highland regions. In fact, our analyses revealed a moderate proportion of parallel adaptation through migration even between the Andes and other highland populations, in contrast to previous studies that appeared to exclude this possibility based on a simple population genetic model (Takuno et al. 2015). This finding is consistent with a recent study which showed that some traits, for example flowering time, may not demonstrate a GxE effect on fitness (Gates et al. 2019), allowing adaptive alleles to move between regions through a matrix of habitat (e.g., the lowlands) where they do not clearly confer an adaptive advantage.

The extensive parallelism in maize highland adaptation has likely been affected by multiple factors and is consistent with expectations based on population genetic theory. First, a high level of genetic diversity in the species provides substantial standing genetic variation and potentially beneficial mutations; indeed, Zhao and Begun (2017) showed that the more genetically diverse species *Drosophila hydei* contained more adaptive alleles than *D. melanogaster*. Second, the recent timing of expansion and divergence in maize has likely affected the extent of parallelism. The probability of parallelism will decrease when populations diverge over a greater period of time, as less shared standing variation would be expected to be the source of adaptation. Increased divergence time would also allow for novel beneficial mutations to arise, potentially in distinct genes and pathways (Conte et al. 2012). For example, Preite et al. (2019) demonstrated a moderate level of molecular parallelism in adaptation to calamine metalliferous soils within *Arabidopsis* species, but less parallelism between species. In addition, a recent study (Bohutínská et al. 2020) discovered that the degree of molecular parallelism decreases with increasing divergence between lineages in alpine *Arabidopsis* species. Third, parallel adaptation from standing genetic variation is unlikely at a large number of loci (MacPherson and Nuismer 2017), highlighting the important role of physiological constraint—which limits the number of loci at which variation can affect a trait—in determining patterns of convergence. Fourth, population genetic modeling predicts that traits with pleiotropic constraint are more likely to demonstrate molecular parallelism (Yeaman 2015). Despite the quantitative nature of flowering time in maize, the C_{χ^2} statistic still showed evidence of genetic constraint and convergence, even after limiting our null to genes in the known flowering time pathway. Similarly, a recent study found molecular parallelism for cold tolerance between distantly related species (lodgepole pine and interior
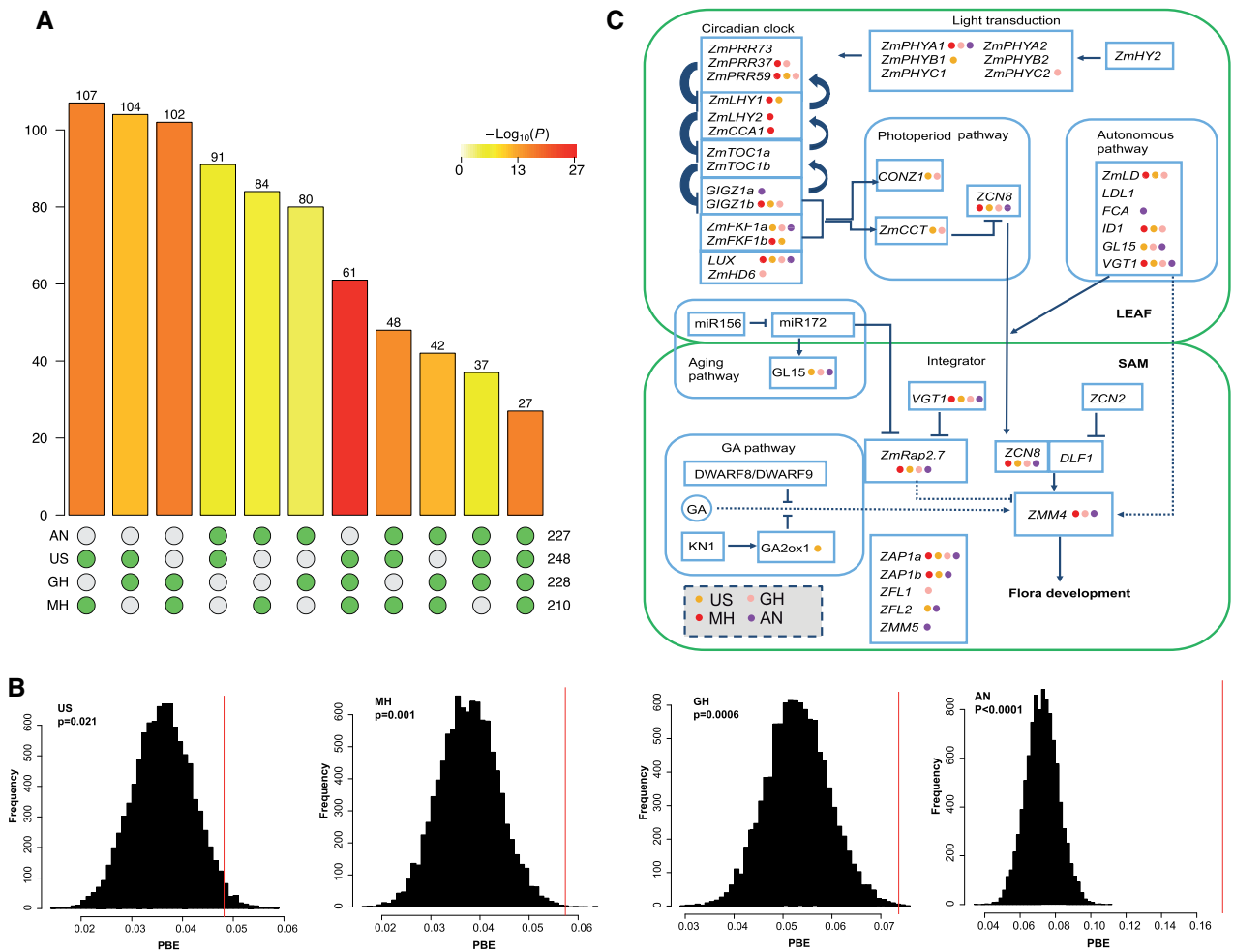
FIG. 5. Convergent adaptation in the flowering time pathway. (A) Intersection of flowering time outlier genes among four highland populations. (B) Distribution of mean *PBE* values of SNPs located within and in the 10-kb flanking regions of core flowering time genes (Dong et al. 2012) (the red line) against a background of all genes (the black bars). (C) Diagram showing selected genes in the flowering time network. Colored dots indicate the population(s) in which selection was detected.

spruce), indicating certain key genes playing crucial roles in cold adaptation (Yeaman et al. 2016). Finally, theory predicts that gene flow may have contrasting impacts on parallel adaptation. Gene flow between differentially adapted populations may introduce maladaptive alleles, counteracting the effects of local selection (Yeaman 2015). For example, Holliday et al. (2016) found islands of divergence in *Populus trichocarpa* populations across altitudinal clines and proposed that coadapted genes in strong linkage may buffer against genetic introgression from maladapted individuals. Our results, consistent with earlier studies (Hufford et al. 2013; Takuno et al. 2015), suggest limited gene flow into the highlands from maladapted lowland maize. Instead, we observe, as a driver of adaptation and parallelism, an important role for adaptive introgression into maize from a highland wild relative and spread of these alleles through gene flow to nearby highland regions. Similarly, adaptive introgression at the EPAS1 locus has been shown to underlie convergent highland adaptation in both humans (Huerta-Sánchez et al. 2014) and dogs (Li et al. 2014).

In summary, the emerging story of parallel highland adaptation in maize is consistent with existing theory. Gene flow is common across maize populations but does not appear to swamp locally adaptive alleles. Rather, highland adaptation appears to have initially been facilitated in the Central Plateau of Mexico by introgression from the locally adapted wild species, *mexicana* (Hufford et al. 2013; Calfee et al. 2021; Rodriguez-Zapata et al. 2021). Subsequently, gene flow from the Mexican highlands appears to have contributed to highland adaptation in other regions of the Americas (Wang et al. 2017), consistent with our finding of migration as a major source of molecular parallelism. Novel alleles from *mexicana* augmented the extensive standing variation already found within the species (Hufford et al. 2012), providing a pool of adaptive variants that appear to have independently risen in frequency in multiple highland regions.

Similar adaptive processes (e.g., gene flow with newly encountered, locally adapted populations and species, repeated selection on standing variation in independent but similar habitats) likely occurred as other crops expanded from their

centers of origin (Janzen et al. 2019). However, crops that were domesticated from wild species with small effective population sizes, with more pronounced domestication bottlenecks, or without widespread, locally adapted congeners may have lacked the adaptive potential necessary to achieve the broad and varied distribution of maize. The success of invasive species has also been linked to hybridization (Heim et al. 2020) and high levels of standing variation (Mimura et al. 2013). A thorough understanding of adaptive processes during rapid expansion across diverse habitats may therefore inform invasive species mitigation strategies, assisted migration in the face of climate change, and crop breeding for tolerance of extreme environmental conditions. Highland adaptation in maize is a clear example in which adaptation has drawn substantially from a shared pool of variants, with the same alleles, genes, and pathways contributing to the evolutionary success of the species across continents and millennia.

Materials and Methods

Samples and Data
Samples and SNPs were from our previous publication (Wang et al. 2017) (supplementary table S1, Supplementary Material online). SNPs were removed when located in centromeric regions, and later filtered to retain only those with sequencing depth greater than $10\times$ and minor allele frequency greater than 0.05 (supplementary fig. S13, Supplementary Material online). SNPs located within known large inversions and *mexicana* introgression regions (primarily in low-recombination regions such as pericentromeres) were removed from analyses due to nonindependence of SNPs within these regions and their enrichment for highly differentiated loci. The filtered SNP data set was shared through the CyVerse platform (https://de.cyverse.org/dl/d/54685BD0-ADCF-4796-8634-12388B097060/wang2020genotype.tar.gz; last accessed April 29, 2021). Nineteen environmental variables were obtained from the WorldClim data set (http://www.worldclim.org/; last accessed April 29, 2021) and were utilized to perform a PCA with the vegan package in R (Oksanen et al. 2016). We employed variant effect predictor from Ensembl (McLaren et al. 2016) to further annotate genomic location and functional consequences of SNPs.

PBS and PBE Calculation
Compared with the commonly used statistic to estimate levels of population differentiation—F_{ST}, *PBS* calculates allele frequency change specific to one focal population by adding pairwise F_{ST} values involving a third "outgroup" population (Yi et al. 2010).

$$PBS = \frac{T^{HL_1} + T^{HL_2} - T^{L_1L_2}}{2} \qquad (1)$$

$$T = -log(1 - F_{ST}) \qquad (2)$$

PBE is a recently described derivative of *PBS*, which overcomes the limitation of *PBS* values being high when all populations have long branches. *PBE* instead quantifies the

difference of the observed and the expected *PBS* values (Yassin et al. 2016).

$$PBE = PBS_{obs} - PBS_{exp} = PBS - \left[T^{L_1L_2} * \frac{PBS_{med}}{T^{L_1L_2}_{med}} \right] \qquad (3)$$

Here, $T^{L_1L_2}$ quantifies genetic differentiation of the two nonfocal populations. *PBE* is expected to be strongly positive when selection acts specifically on the focal population.

We calculated *PBE* for filtered SNPs among one highland population, its corresponding lowland population (the Mexican lowland population was paired with the Mesoamerican highland populations; the South American lowland population was paired with the Andes population) and the *parviglumis* Palmar Chico population (supplementary fig. S13, Supplementary Material online).

Pairwise Weir–Cockerham F_{ST} values were calculated in vcfTools (Danecek et al. 2011) and custom R scripts were used to compute *PBS* and *PBE* values of the focal highland population. SNPs with *PBE* values higher than the 95% quantile of its distribution were regarded as outliers and qualitative results were confirmed in the 99% quantile. The R package "SuperExactTest" (Wang et al. 2018) was utilized to evaluate if the overlap of outlier SNPs between pairwise highland populations was enriched. In order to address false detection of overlap driven by linkage disequilibrium among loci, we removed loci with an r^2 greater than 0.2 within 10kb windows and re-evaluated the significance of overlap among highland populations.

In addition, we also evaluated if the same allele is elevated to high frequency among outlier SNPs between pairwise highland populations (supplementary fig. S14, Supplementary Material online). If there is no allele frequency change between the highland and its lowland counterpart, we categorized the SNPs as "nondirectional." When a SNP was nondirectional in both highland populations being compared, it was removed from the analyses. When the same allele was at elevated frequency in both highland populations compared with their lowland counterparts, the SNP was categorized as "codirectional". When different alleles were at high frequency across two highland populations, the SNP was categorized as "antidirectional". We then compared the proportion of co-directional SNPs between the outlier and neutral SNPs.

In order to exclude the bias of genetic drift on the directionality of changes in allele frequency, we approximated the two-dimensional site frequency spectrum (2dSFS) of outlier SNPs using a subset of the neutral SNP set. We divided the nonreference allele frequency in Mexican Lowland maize and *parviglumis* into ten equal bins in both the outlier and neutral SNP set. Then we matched neutral SNPs with the same ancestral 2dSFS (representing the combination of allele frequencies in both the Mexican Lowland and *parviglumis* populations) with the outlier SNP set and sampled the same amount of SNPs in each allele frequency bin. Last, we

tabulated the directionality of the randomly sampled neutral SNPs to check how ancestral allele frequency influenced the proportion of codirectional SNPs.

Composite Likelihood Calculations to Distinguish the Mode of Parallelism

Following the method developed and applied in Lee and Coop (2017) and Tittes (2020), we calculated the composite likelihood under four models for the outlier regions shared among all four highland populations, after thinning for linkage. The four models tested were as follows: 1) independent mutations of the beneficial allele in four highland populations, 2) a single origin of the beneficial allele in one single highland population and spread via gene flow into the other population during the sweep, 3) standing variation present in the ancestor of the four highland populations, 4) gene flow contributing novel standing variation which, t generations later, is selected. The details of these four models can be found in Lee and Coop (2017). For each outlier, we obtained SNPs in the surrounding 20-kb window and calculated the composite likelihood under each of the four models for this data set. We compared composite likelihoods under the models for each outlier and obtained maximum composite-likelihood-parameter estimates. Similarly, we also evaluated the composite likelihood values under four models for each set of dually selected SNPs in pairs of highland populations.

Constraint of the Adaptation Target Size

Following the method in Yeaman et al. (2018), we first based our analyses on a null hypothesis that all genes contribute equally to adaptation. Second, we set up a new null hypothesis that only the selected genes (those selected in at least one highland population) could contribute to adaptive traits. With those two different null hypotheses, we calculated C_{χ^2} and the corresponding P values and compared the differences.

Polygenic Adaptation for Quantitative Traits

We followed the methods developed by Berg and Coop (2014) and Josephs (2018) to detect polygenic adaptation in the highland populations. Briefly, these methods detect adaptive divergence for a trait by 1) finding loci associated with that trait in a GWAS, 2) summarizing the allele frequencies and effect sizes at these loci in the populations of interest using a polygenic score, and 3) testing to see if the association between these polygenic scores and an environmental character (in this case, elevation) are greater than could be explained by drift.

To identify loci associated with traits that could be under selection, we conducted GWAS using a maize panel developed for GWAS, often referred to as "the 282" or "the Major Goodman panel" (Flint-Garcia et al. 2005). SNPs for 263 individuals in the Major Goodman panel were called from whole genome sequencing data from Bukowski et al. (2017), removing individuals with genotype calls for <70% of polymorphic sites. Of all the SNPs called in Bukowski et al. (2017), we only used those that were also polymorphic in the maize landraces and that had a MAF > 0.01 and were missing data for <5% of

individuals, leaving us with approximately 5 million SNPs per test. We used trait measurements made in three short-day common garden experiments, from Florida in 2006 and 2007 and Puerto Rico in 2006, from Hung et al. (2012) where there were data available for at least 80% of the 263 individuals, leaving us with 29 trait-environment combinations. We conducted GWAS using GEMMA with default parameters (Zhou and Stephens 2012), with a standardized kinship matrix to control for population structure.

We used the GWAS hits to generate polygenic scores for each individual in the landrace panel using all loci with a P value < 0.1 that had been pruned down to the strongest hit per 1 cM window using a linkage map from Ogut et al. (2015), leaving us with an average of 216 SNPs per trait (range = 147–289). If there are M loci associated with a trait, each with effect size α_m, p_{im} is the allele frequency of the mth allele in line i, which will either be 0, 0.5, or 1, then we can calculate the polygenic score for the ith line as Z_i where

$$Z_i = 2 \sum_{m=1}^{M} \alpha_m p_{im} \qquad (4)$$

Shared population structure between the GWAS panel and the test panel can lead to false positive signals of selection (Josephs 2018). To test for shared structure between the GWAS panel used here and the landraces, we constructed a joint kinship matrix between all samples in both panels, following the same methods used in Josephs (2018). Plotting the first two eigenvectors of this joint kinship matrix (the principal components) shows that there is likely shared population structure between the landraces and the GWAS panel (supplementary fig. S15A, Supplementary Material online). We constructed a conditional kinship matrix that estimates the amount of relatedness between landraces conditional on relatedness captured in the GWAS panel, and found that this new conditional matrix explains less variation than the original matrix made from landraces alone (supplementary fig. S15B, Supplementary Material online). This finding is further evidence that there is shared population structure between the two panels.

In light of this shared structure, we conducted a conditional test, following Josephs (2018). Let $\vec{Z}_1$ be the vector of Z_i for the 31 highland and lowland landraces discussed in this paper and $\vec{Z}_2$ is the vector of Z_i for the individuals used in the GWAS. We model the combined vector of polygenic scores in both panels as

$$\begin{pmatrix} \vec{Z}_1 \\ \vec{Z}_2 \end{pmatrix} \sim MVN \left(\begin{pmatrix} \mu \\ \mu \end{pmatrix}, V_A \begin{pmatrix} K_{11} & K_{12} \\ K_{12}^T & K_{22} \end{pmatrix} \right). \qquad (5)$$

where, μ is the mean of the combined vector $[X_1, X_2]$, K_{11} and K_{22} are the kinship matrices of the genotyping and GWAS panels, and K_{12} is the set of relatedness coefficients between lines in the genotyping panel (rows) and GWAS panel (columns).

The conditional multivariate null model for our polygenic scores in the landraces conditional on the GWAS panel is then

$$\vec{Z}_1 | \vec{Z}_2 \sim MVN(\vec{\mu}', 2V_A K'), \qquad (6)$$

where $\vec{\mu}'$ is a vector of conditional means with an entry for each sample in the genotyping panel:

$$\vec{\mu}' = \mu + K_{12} K_{22}^{-1} (\vec{Z}_2 - \mu) \qquad (7)$$

and K' is the relatedness matrix for the genotyping panel conditional on the matrix of the GWAS panel,

$$K' = K_{11} - K_{12} K_{22}^{-1} K_{12}^T. \qquad (8)$$

To test for selection, we calculate the difference between observed polygenic scores and conditional means as $\vec{Z} = \vec{Z}_1 - \vec{\mu}'$. Higher values of $\vec{Z}$ indicate that the observed polygenic score in a landrace individual is greater than would be expected based on relatedness between that landrace individual and individuals in the GWAS panel.

After calculating $\vec{Z}$ for 29 environment-trait combinations, we test for a correlation between $\vec{Z}$ and the elevation of origin of each landrace beyond what would be expected due to neutral drift using the methods described in Berg and Coop (2014). We transformed both $\vec{Z}$ and the mean-centered vector of elevations for each landrace by the Cholesky decomposition of the kinship matrix (C) using the following equation (shown here for $\vec{Z}$)

$$\vec{X} = \frac{1}{\sqrt{2V_A}} C^{-1} (\vec{Z}') \qquad (9)$$

We estimate V_A using the allele frequencies and effect sizes of the GWAS loci, using the following formula. If $\bar{p}_m$ is the allele frequency of the mth locus across all individuals and α_m is the effect size of that locus,

$$V_A = \sum_{m=1}^{M} \alpha_m^2 \bar{p}_m (1 - \bar{p}_m) \qquad (10)$$

We then test for a linear relationship between $\vec{X}$ and the similarly transformed vector of elevations using the lm function in R (R Core Team 2018) and control for the 29 tests done using Qvalue (Storey and Tibshirani 2003).

Supplementary Material

Supplementary data are available at *Molecular Biology and Evolution* online.

Acknowledgments

References

Alves JM, Carneiro M, Cheng JY, de Matos AL, Rahman MM, Loog L, Campos PF, Wales N, Eriksson A, Manica A, et al. 2019. Parallel adaptation of rabbit populations to myxoma virus. *Science* 363(6433):1319–1326.

Berg JJ, Coop G. 2014. A population genetic signal of polygenic adaptation. *PLoS Genet.* 10(8):e1004412.

Besnard G, Muasya AM, Russier F, Roalson EH, Salamin N, Christin P-A. 2009. Phylogenomics of c4 photosynthesis in sedges (cyperaceae): multiple appearances and genetic convergence. *Mol Biol Evol.* 26(8):1909–1919.

Bohutinska M, Vlcek J, Yair S, Leanen B, Konecna V, Fracassetti M, Slotte T, Kolar F. 2020. Genomic basis of parallel adaptation varies with divergence in arabidopsis and its relatives. *bioRxiv*.

Brandenburg J-T, Mary-Huard T, Rigaill G, Hearne SJ, Corti H, Joets J, Vitte C, Charcosset A, Nicolas SD, Tenaillon MI. 2017. Independent introductions and admixtures have contributed to adaptation of European maize and its American counterparts. *PLoS Genet.* 13(3):e1006666.

Buckler ES, Holland JB, Bradbury PJ, Acharya CB, Brown PJ, Browne C, Ersoz E, Flint-Garcia S, Garcia A, Glaubitz JC, et al. 2009. The genetic architecture of maize flowering time. *Science* 325(5941):714–718.

Bukowski R, Guo X, Lu Y, Zou C, He B, Rong Z, Wang B, Xu D, Yang B, Xie C, et al. 2017. Construction of the third-generation *Zea mays* haplotype map. *GigaScience* 7(4):gix134.

Bush M, Correa-Metrio A, McMichael C, Sully S, Shadik C, Valencia B, Guilderson T, Steinitz-Kannan M, Overpeck J. 2016. A 6900-year history of landscape modification by humans in lowland amazonia. *Quat Sci Rev.* 141:52–64.

Calfee E, Gates D, Lorant A, Perkins MT, Coop G, Ross-Ibarra J. 2021. Selective sorting of ancestral introgression in maize and teosinte along an elevational cline. *bioRxiv*.

Christin P-A, Salamin N, Savolainen V, Duvall MR, Besnard G. 2007. C4 photosynthesis evolved in grasses via parallel adaptive genetic changes. *Curr Biol.* 17(14):1241–1247.

Conte GL, Arnegard ME, Peichel CL, Schluter D. 2012. The probability of genetic parallelism and convergence in natural populations. *P Roy Soc B Biol Sci.* 279(1749):5039–5047.

da Fonseca RR, Smith BD, Wales N, Cappellini E, Skoglund P, Fumagalli M, Samaniego JA, Carøe C, Ávila-Arcos MC, Hufnagel DE, et al. 2015. The origin and evolution of maize in the American southwest. *Nat Plants.* 1:14003.

Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, et al. 2011. The variant call format and vcftools. *Bioinformatics* 27(15):2156–2158.

Dasmahapatra KK, Walters JR, Briscoe AD, Davey JW, Whibley A, Nadeau NJ, Zimin AV, Hughes DS, Ferguson LC, Martin SH, et al. 2012. Butterfly genome reveals promiscuous exchange of mimicry adaptations among species. *Nature* 487(7405):94.

Davies KT, Bennett NC, Faulkes CG, Rossiter SJ. 2018. Limited evidence for parallel molecular adaptations associated with the subterranean niche in mammals: a comparative study of three superorders. *Mol Biol Evol.* 35(10):2544–2559.

Dickau R, Ranere AJ, Cooke RG. 2007. Starch grain evidence for the preceramic dispersals of maize and root crops into tropical dry and humid forests of panama. *Proc Natl Acad Sci U S A.* 104(9):3651–3656.

Dobler S, Dalla S, Wagschal V, Agrawal AA. 2012. Community-wide convergent evolution in insect adaptation to toxic cardenolides by substitutions in the na, k-atpase. *Proc Natl Acad Sci U S A.* 109(32):13040–13045.

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

Eagles H, Lothrop J. 1994. Highland maize from central mexico-its origin, characteristics, and use in breeding programs. *Crop Sci.* 34(1):11–19.

Edelman NB, Frandsen PB, Miyagi M, Clavijo B, Davey J, Dikow RB, García-Accinelli G, Van Belleghem SM, Patterson N, Neafsey DE, et al. 2019. Genomic architecture and introgression shape a butterfly radiation. *Science* 366(6465):594–599.

Emerson TE, Hedman KM, Simon ML, Fort MA, Witt KE. 2020. Isotopic confirmation of the timing and intensity of maize consumption in greater cahokia. *Am Antiq.* 85(2):241–262.

Flint-Garcia SA, Thuillet A-C, Yu J, Pressoir G, Romero SM, Mitchell SE, Doebley J, Kresovich S, Goodman MM, Buckler ES. 2005. Maize association population: a high-resolution platform for quantitative trait locus dissection. *Plant J.* 44(6):1054–1064.

Gates DJ, Runcie D, Janzen GM, Navarro AR, Willcox M, Sonder K, Snodgrass S, Rodríguez-Zapata F, Sawers RJ, Rellán-Álvarez R, et al. 2019. Single-gene resolution of locally adaptive genetic variation in Mexican maize. *bioRxiv.* 706739.

Grobman A, Bonavia D, Dillehay TD, Piperno DR, Iriarte J, Holst I. 2012. Preceramic maize from paredones and huaca prieta, peru. *Proc Natl Acad Sci U S A.* 109(5):1755–1759.

Guo L, Wang X, Zhao M, Huang C, Li C, Li D, Yang CJ, York AM, Xue W, Xu G, et al. 2018. Stepwise cis-regulatory changes in zcn8 contribute to maize flowering-time adaptation. *Curr Biol.* 28(18):3005–3015.

Hall SA. 2010. Early maize pollen from Chaco Canyon, New Mexico, USA. *Palynology* 34(1):125–137.

Heim KC, McMahon TE, Ertel BD, Koel TM. 2020. Leveraging public harvest to reduce invasive hybridization in yellowstone national park: field identification and harvest of cutthroat x rainbow trout hybrids. *Biol Invasions.* 22:2685–2698.

Holliday JA, Zhou L, Bawa R, Zhang M, Oubida RW. 2016. Evidence for extensive parallelism but divergent genomic architecture of adaptation along altitudinal and latitudinal gradients in *Populus trichocarpa.* *New Phytol.* 209(3):1240–1251.

Huerta-Sánchez E, Jin X, Bianba Z, Peter BM, Vinckenbosch N, Liang Y, Yi X, He M, Somel M, Ni P, et al. 2014. Altitude adaptation in Tibetans caused by introgression of Denisovan-like DNA. *Nature* 512(7513):194–197.

Hufford MB, Xu X, Van Heerwaarden J, Pyhäjärvi T, Chia J-M, Cartwright RA, Elshire RJ, Glaubitz JC, Guill KE, Kaeppler SM, et al. 2012. Comparative population genomics of maize domestication and improvement. *Nat Genet.* 44:808–811.

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

Hung H, Browne C, Guill K, Coles N, Eller M, Garcia A, Lepak N, Melia-Hancock S, Oropeza-Rosas M, Salvo S, et al. 2012. The relationship between parental genetic or phenotypic divergence and progeny variation in the maize nested association mapping population. *Heredity* 108(5):490–499.

Janzen GM, Wang L, Hufford MB. 2019. The extent of adaptive wild introgression in crops. *New Phytol.* 221(3):1279–1288.

Jiang C, Edmeades G, Armstead I, Lafitte H, Hayward M, Hoisington D. 1999. Genetic analysis of adaptation differences between highland and lowland tropical maize using molecular markers. *Theor. Appl. Genet.* 99(7-8):1106–1119.

Josephs EB. 2018. Determining the evolutionary forces shaping G x E. *New Phytol.* 219(1):31–36.

Josephs EB, Berg JJ, Ross-Ibarra J, Coop G. 2019. Detecting adaptive differentiation in structured populations with genomic data and common gardens. *Genetics* 211(3):989–1004.

Kistler L, Maezumi SY, Gregorio de Souza J, Przelomska NAS, Malaquias Costa F, Smith O, Loiselle H, Ramos-Madrigal J, Wales N, Ribeiro ER, et al. 2018. Multiproxy evidence highlights a complex evolutionary legacy of maize in South America. *Science* 362(6420):1309–1313.

Korner C. 2007. The use of 'altitude' in ecological research. *Trends Ecol Evol.* 22(11):569–574.

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

Lee KM, Coop G. 2017. Distinguishing among modes of convergent adaptation using population genomic data. *Genetics* 207(4):1591–1619.

Li Y, Wu D-D, Boyko AR, Wang G-D, Wu S-F, Irwin DM, Zhang Y-P. 2014. Population variation revealed high-altitude adaptation of Tibetan mastiffs. *Mol Biol Evol.* 31(5):1200–1205.

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

Lomax BH, Fraser WT, Harrington G, Blackmore S, Sephton MA, Harris NB. 2012. A novel palaeoaltimetry proxy based on spore and pollen wall chemistry. *Earth Planet Sci Lett.* 353–354:22–28.

MacPherson A, Nuismer S. 2017. The probability of parallel genetic evolution from standing genetic variation. *J Evol Biol.* 30(2):326–337.

Malinsky M, Matschiner M, Svardal H. 2021. Dsuite-fast d-statistics and related admixture evidence from vcf files. *Mol Ecol Resour.* 21(2):584–595.

Matsuoka Y, Vigouroux Y, Goodman MM, Sanchez G J, Buckler E, Doebley J. 2002. A single domestication for maize shown by multilocus microsatellite genotyping. *Proc Natl Acad Sci U S A.* 99:6080–6084.

McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GR, Thormann A, Flicek P, Cunningham F. 2016. The ensembl variant effect predictor. *Genome Biol.* 17(1):122.

Merrill WL, Hard RJ, Mabry JB, Fritz GJ, Adams KR, Roney JR, MacWilliams AC. 2009. The diffusion of maize to the southwestern united states and its impact. *Proc Natl Acad Sci U S A.* 106(50):21019–21026.

Mimura M, Ono K, Goka K, Hara T. 2013. Standing variation boosted by multiple sources of introduction contributes to the success of the introduced species, lotus corniculatus. *Biol Invasions.* 15(12):2743–2754.

Navarro JAR, Willcox M, Burgueño J, Romay C, Swarts K, Trachsel S, Preciado E, Terron A, Delgado HV, Vidal V, et al. 2017. A study of allelic diversity underlying flowering-time adaptation in maize landraces. *Nat Genet.* 49(3):476–480.

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

Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, McGlinn D, Minchin PR, O'Hara RB, Simpson GL, Solymos P, et al. 2016. vegan: community ecology package. R package version. 2.4-0. Available from: https://CRAN.R-project.org/package=vegan. Accessed April 29, 2021.

Orr HA. 2005. The probability of parallel adaptation. *Evolution* 59(1):216–220.

Piperno DR, Flannery KV. 2001. The earliest archaeological maize (*Zea mays* L.) from highland Mexico: new accelerator mass spectrometry dates and their implications. *Proc Natl Acad Sci U S A.* 98(4):2101–2103.

Piperno DR, Moreno JE, Iriarte J, Holst I, Lachniet M, Jones JG, Ranere AJ, Castanzo R. 2007. Late pleistocene and holocene environmental history of the Iguala valley, central balsas watershed of Mexico. *Proc Natl Acad Sci U S A.* 104(29):11874–11881.

Piperno DR, Ranere AJ, Holst I, Iriarte J, Dickau R. 2009. Starch grain and phytolith evidence for early ninth millennium bp maize from the Central Balsas River Valley, Mexico. *Proc Natl Acad Sci U S A.* 106(13):5019–5024.

Pool JE, Braun DT, Lack JB. 2017. Parallel evolution of cold tolerance within *Drosophila melanogaster.* *Mol Biol Evol.* 34(2):349–360.

Preite V, Sailer C, Syllwasschy L, Bray S, Ahmadi H, Krämer U, Yant L. 2019. Convergent evolution in *Arabidopsis halleri* and *Arabidopsis arenosa* on calamine metalliferous soils. *Philos Trans R Soc B.* 374(1777):20180243.

R Core Team. 2018. R foundation for statistical computing R: a language and environment for statistical computing. Vienna, Austria.

Rodriguez J, Sánchez G, Baltazar B, De la Cruz L, Santacruz-Ruvalcaba F, Ron P, Schoper J. 2006. Characterization of floral morphology and synchrony among Zea species in Mexico. *Maydica*. 51(2):383–398.

Rodriguez-Zapata F, Barnes AC, Blocker-Juarez KA, Gates D, Kur A, Wang L, Janzen GM, Jensen S, Estevez-Palmas JM, Crow T, et al. 2021. Teosinte introgression modulates phosphatidylcholine levels and induces early maize flowering time. bioRxiv. doi: 10.1101/2021.01.25.426574.

Rosenblum EB, Parent CE, Brandt EE. 2014. The molecular basis of phenotypic convergence. *Annu Rev Ecol Evol Syst*. 45(1):203–226.

Ross-Ibarra J, Piperno D. 2020. Maize moving. figshare. Figure. Available from: 10.6084/m9.figshare.12781307.v1

Stephen Athens J, Ward JV, Pearsall DM, Chandler-Ezell K, Blinn DW, Morrison AE. 2016. Early prehistoric maize in northern highland Ecuador. *Latin Am Antiq*. 27(1):3–21.

Storey JD, Tibshirani R. 2003. Statistical significance for genomewide studies. *Proc Natl Acad Sci U S A*. 100(16):9440–9445.

Storz JF. 2016. Causes of molecular convergence and parallelism in protein evolution. *Nat Rev Genet*. 17(4):239.

Takuno S, Ralph P, Swarts K, Elshire RJ, Glaubitz JC, Buckler ES, Hufford MB, Ross-Ibarra J. 2015. Independent molecular basis of convergent highland adaptation in maize. *Genetics* 200(4):1297–1312.

Tittes S. 2020. rdmc: an open source r package implementing convergent adaptation models of Lee and Coop (2017). *Genes Genomes Genetics*. 10(9):3041–3046.

van Heerwaarden J, Doebley J, Briggs WH, Glaubitz JC, Goodman MM, de Jesus Sanchez Gonzalez J, Ross-Ibarra J. 2011. Genetic signals of origin, spread, and introgression in a large sample of maize landraces. *Proc Natl Acad Sci U S A*. 108(3):1088–1092.

Wang B, Lin Z, Li X, Zhao Y, Zhao B, Wu G, Ma X, Wang H, Xie Y, Li Q, et al. 2020. Genome-wide selection and genetic improvement during modern maize breeding. *Nat Genet*. 52:565–571.

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

Wang M, Zhao Y, Zhang B. 2018. SuperExactTest: exact test and visualization of multi-set intersections. R package version 1.0.0. Available from: https://CRAN.R-project.org/package=SuperExactTest. Accessed April 29, 2021.

Yang L, Wang H-N, Hou X-H, Zou Y-P, Han T-S, Niu X-M, Zhang J, Zhao Z, Todesco M, Balasubramanian S, et al. 2018. Parallel evolution of common allelic variants confers flowering diversity in *Capsella rubella*. *Plant Cell*. 30(6):1322–1336.

Yassin A, Debat V, Bastide H, Gidaszewski N, David JR, Pool JE. 2016. Recurrent specialization on a toxic fruit in an island drosophila population. *Proc Natl Acad Sci U S A*. 113(17):4771–4776.

Yeaman S. 2015. Local adaptation by alleles of small effect. *Am Nat*. 186(S1):S74–S89.

Yeaman S, Gerstein AC, Hodgins KA, Whitlock MC. 2018. Quantifying how constraints limit the diversity of viable routes to adaptation. *PLoS Genet*. 14:e1007717.

Yeaman S, Hodgins KA, Lotterhos KE, Suren H, Nadeau S, Degner JC, Nurkowski KA, Smets P, Wang T, Gray LK, et al. 2016. Convergent local adaptation to climate in distantly related conifers. *Science* 353(6306):1431–1433.

Yi X, Liang Y, Huerta-Sanchez E, Jin X, Cuo ZXP, Pool JE, Xu X, Jiang H, Vinckenbosch N, Korneliussen TS, et al. 2010. Sequencing of 50 human exomes reveals adaptation to high altitude. *Science* 329(5987):75–78.

Zhao L, Begun DJ. 2017. Genomics of parallel adaptation at two timescales in Drosophila. *PLoS Genet*. 13(10):e1007016.

Zhen Y, Aardema ML, Medina EM, Schumer M, Andolfatto P. 2012. Parallel molecular evolution in an herbivore community. *Science* 337(6102):1634–1637.

Zhou X, Stephens M. 2012. Genome-wide efficient mixed-model analysis for association studies. *Nat Genet*. 44(7):821–824.