Complex Patterns of Local Adaptation in Teosinte

Tanja Pyhäjärvi[1,2], Matthew B. Hufford[1], Sofiane Mezmouk[1], and Jeffrey Ross-Ibarra[1,3,4,*]

[1]Department of Plant Sciences, University of California, Davis

[2]Department of Biology, University of Oulu, Oulu, Finland

[3]Center for Population Biology, University of California, Davis

[4]The Genome Center, University of California, Davis

***Corresponding author**: E-mail: rossibarra@ucdavis.edu.

Accepted: July 19, 2013

Data deposition: SNP data, map positions, and annotations deposited in the Dryad repository: doi:10.5061/dryad.8m648.

Abstract

Populations of widely distributed species encounter and must adapt to local environmental conditions. However, comprehensive characterization of the genetic basis of adaptation is demanding, requiring genome-wide genotype data, multiple sampled populations, and an understanding of population structure and potential selection pressures. Here, we used single-nucleotide polymorphism genotyping and data on numerous environmental variables to describe the genetic basis of local adaptation in 21 populations of teosinte, the wild ancestor of maize. We found complex hierarchical genetic structure created by altitude, dispersal events, and admixture among subspecies, which complicated identification of locally beneficial alleles. Patterns of linkage disequilibrium revealed four large putative inversion polymorphisms showing clinal patterns of frequency. Population differentiation and environmental correlations suggest that both inversions and intergenic polymorphisms are involved in local adaptation.

Key words: *Zea mays*, *parviglumis*, *mexicana*, inversion, population structure, admixture.

Introduction

One of the enduring goals of evolutionary genetics is to understand the genomic and geographic extent of adaptation in natural populations. Adaptive differences among closely related contemporary populations are prime examples of the impact of natural selection on genetic polymorphisms. While many examples of adaptation in nature are known (Ford and Ford 1930; Clausen et al. 1940; Endler 1980; Grant and Grant 2002), our theoretical understanding (Lewontin and Krakauer 1973; Feder and Nosil 2010; Le Corre and Kremer 2012) of the genomic signatures of adaptation has so far outpaced natural studies, and a number of outstanding questions remain. For example, what kinds of traits are adaptive? What are the prevailing forms (i.e., clinal, local) of adaptation in nature? What parts of the genome and what kinds of loci are involved?

One approach to link adaptation to genetic diversity is the reverse ecology (Li et al. 2008) strategy of identifying genomic regions that have been the target of natural selection in order to understand ecologically important variation. The advantage of this approach is that it is not dependent on experimental setup or identification of relevant, measurable traits. This approach uses a number of statistical and population genetic methods to identify putatively selected loci. Such loci are usually expected to show higher genetic differentiation among populations and lower genetic diversity within populations (e.g., Lewontin and Krakauer 1973; Storz 2005), and should be correlated with environmental variation that is directly or indirectly causing the selection pressure (Coop et al. 2010; Eckert et al. 2010a). Several related methods to identify loci that have exceptionally high differentiation are available (Beaumont and Nichols 1996; Beaumont and Balding 2004; Foll and Gaggiotti 2008). These are typically based on identifying F_{ST} outliers by generating null distributions of F_{ST} from a simple neutral model. Results of such methods may depend strongly on the accuracy of the null model: hierarchical population structure, for example, can lead to a large number of false positives (Excoffier et al. 2009). An alternative approach involves identifying associations between environmental variables and genetic polymorphisms while controlling for population structure. This method has been successfully applied to

identify adaptively important loci among populations in humans (Coop et al. 2010; Hancock et al. 2011a) and *Arabidopsis thaliana* (Fournier-Level et al. 2011; Hancock et al. 2011b).

For plants, adaptation to local conditions is especially important because, as sedentary organisms, they cannot readily migrate to more benign conditions. The advent of high-throughput genotyping has begun to enable researchers to answer some of the questions related to the genetic basis of local adaptation in plants. For example, recent studies in *Arabidopsis* have shown geographically localized fitness effects (Fournier-Level et al. 2011), enrichment for non-synonymous variants at adaptive loci (Hancock et al. 2011b), and adaptation across a number of environmental variables (e.g., soil type, temperature, and precipitation) (Turner et al. 2010; Fournier-Level et al. 2011). However, although both global (Platt et al. 2010) and local (Bomblies et al. 2010) population structure in *Arabidopsis* can be complex, little is known of how this has impacted studies of local adaptation. Moreover, *Arabidopsis* has an unusually small and relatively simple genome among plants (Meyerowitz and Pruitt 1985), and it is not yet clear how generally applicable results from this model species are to other taxa.

The wild relatives of domesticated maize offer an excellent opportunity to investigate questions of local adaptation in natural populations (Hufford, Bilinski, et al. 2012). In addition to the well-known domesticated maize, *Zea mays* ssp. *mays*, the species *Zea mays* includes three wild subspecies, collectively known as teosintes. Two of these, *Zea mays* ssp. *parviglumis* (hereafter *parviglumis*) and *Zea mays* ssp. *mexicana* (hereafter *mexicana*), are widespread taxa found in natural populations spanning central and southwest Mexico (fig. 1). Their combined range extends across varied environments, from the warmer low elevations of the Balsas River Valley to the cooler high elevations of Mexico's Central Plateau, encompassing 2-fold differences in precipitation and several soil types. Ecological niche modeling suggests that both taxa have moved little in response to climate change since the Last Glacial Maximum (approximately 20,000 base pair) (Hufford, Martínez-Meyer, et al. 2012), and their annual life history and outcrossing mating system have likely facilitated adaptation to local conditions. The teosintes also exhibit complex hierarchical population structure (Fukunaga et al. 2005) and have a genome of relatively average size and complexity among angiosperms (Gregory et al. 2007), thus serving as a valuable counterpoint to results from *Arabidopsis*.

Previous work has investigated the genetic architecture of morphological differences in the teosintes (Lauter et al. 2004; Weber et al. 2007, 2008), and overall patterns of genetic diversity and structure among populations have been detailed using various genetic markers (Smith et al. 1984; Doebley et al. 1987; Fukunaga et al. 2005; Moeller et al. 2007). Our understanding of patterns of variation across the genome, however, has been limited to a handful of accessions or analyses of individual populations (van Heerwaarden et al. 2010; Hufford et al. 2012). Even less is known regarding the genetics of local adaptation, which has only been studied at a few loci in the context of plant immunity (Moeller and Tiffin 2008).

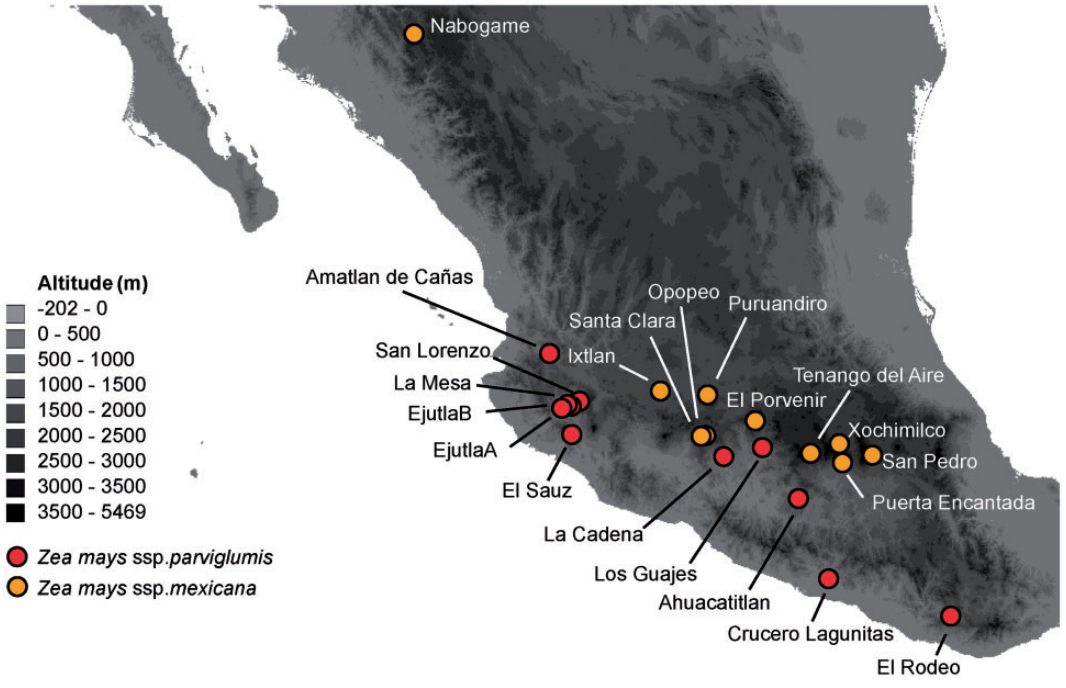

Fig. 1.—Map of sampled *Zea mays* ssp. *parviglumis* and ssp. *mexicana* populations.

Clinal patterns of large inversions in *Drosophila* were one of the first observed adaptive patterns in genetic polymorphisms (Dobzhansky 1970). A major effect of chromosomal inversions is recombination suppression, and early on it was suggested that clinal patterns were a result of selection acting on large co-adapted gene complexes caught by inversions (Wright and Dobzhansky 1946; Dobzhansky 1970). Recent theoretical work has shown that no functional connection among genes within inversions is necessary; for selection to maintain inversion polymorphisms they only need to capture two or more locally beneficial alleles (Kirkpatrick and Barton 2006; Kirkpatrick 2010). Large genomic rearrangements are often observed in maize and its wild relatives (Ting 1964; Kato 1975), but their adaptive significance is generally not known. The single exception to this is an observation of an altitudinal cline and unusual haplotype structure at a large inversion found segregating in multiple *parviglumis* and *mexicana* populations (Fang et al. 2012).

Here we present a detailed population genetic examination of local adaptation across 21 populations of *mexicana* and *parviglumis* (fig. 1) based on single-nucleotide polymorphism (SNP) genotyping. We document complex, hierarchical population structure patterned by elevation, dispersal, and recent admixture. We scan the genome for alleles associated with excess differentiation or strong correlation with environmental gradients and show that population structure complicates our ability to identify putatively adaptive loci. Nonetheless, we identify a number of candidate SNPs, four large, putatively adaptive inversion polymorphisms, and an excess of loci in intergenic regions of the genome. These results differ markedly from previous work in other plant species, suggesting that the genetics of local adaptation may vary considerably among taxa.

Materials and Methods

Sampling

Ten to 12 individuals were sampled from each of 11 populations of *parviglumis* and 10 populations of *mexicana* (fig. 1 and supplementary table S1, Supplementary Material online). Seeds from four populations of *parviglumis* (Ejutla A and B, San Lorenzo, and La Mesa) were sampled by Matthew B. Hufford in 2007, while the remaining seven populations were obtained from nonregenerated bulked seed from accessions in the USDA germplasm collections. All *mexicana* populations were collected in 2008 and kindly provided by Pesach Lubinsky and Norman Ellstrand. Two *mexicana* individuals were removed from subsequent analysis after being identified as recent hybrids with maize in an initial principal component analysis (PCA) that included data from 279 maize inbred lines (Cook et al. 2012). A *Tripsacum dactyloides* (Nei's divergence to Zea = 0.06, Ross-Ibarra et al. 2009) sample provided by Sherry Flint-Garcia was used as an outgroup. The geodetic distances among populations vary from 3 km between Santa Clara and Opopeo to 1,503 km between El Rodeo and Nabogame, and the populations cover an altitudinal range from 590 to 2,609 m above sea level. Previously published genotype data (Cook et al. 2012) and phenotypic data from 278 maize inbred lines (Flint-Garcia et al. 2005) were used for population structure and trait association analyses.

SNP Genotyping

Leaf tips of seedlings at the five-leaf stage were collected, stored at −80 °C overnight, and lyophilized for 48 h. DNA was extracted from homogenized tissue following a modified CTAB protocol (Saghai-Maroof et al. 1984) and quantified using a NanoDrop spectrophotometer (NanoDrop Technologies, Inc., Wilmington, DE, USA) and Wallac VICTOR2 fluorescence plate reader (Perkin-Elmer Life and Analytical Sciences, Torrance, CA) with the Quant-iT™ PicoGreen® dsDNA Assay Kit (Invitrogen, Grand Island, NY, USA).

SNP genotyping was conducted at the UC Davis Genome Center using the MaizeSNP50 BeadChip and Infinium® HD Assay (Illumina, San Diego, CA, USA). SNP genotypes were called using GenomeStudio V2009.1 (Illumina). After dropping SNPs with >10% missing data in either subspecies and manual adjustment of clustering, average call rates (the proportion of successfully called genotypes) per individual were 99% for both *parviglumis* and *mexicana*. Genotyping error estimated from three technical replicates of *parviglumis* was approximately 0.0015% after mismatches caused by missing data were excluded. In total, 43,701 SNPs were called in *parviglumis* and 43,694 in *mexicana*. SNPs within a region on chromosome 8 that showed long regions of deviation from Hardy–Weinberg equilibrium in individual populations and showed discrepancies between the physical and genetic maps in Ganal et al. (2011) were removed. After removing monomorphic and duplicate SNPs and SNPs that did not map or mapped multiply to the maize reference genome (release 5b.60), the final data set consisted of 36,719 SNPs genotyped in both subspecies. SNP data, map positions, and annotations are available at http://datadryad.org/resource/doi:10.5061/dryad.8m648.

Additional, more rigorous filtering was applied in haplotype-based analyses that are especially sensitive to the relationship between physical and genetic map positions. Genetic map coordinates for fastPHASE were obtained from a modified map (Gerke et al. 2013) of the maize Intermated B73 × Mo17 (IBM) population (Ganal et al. 2011). The 3,615 SNPs for which genetic and physical map disagreed or information was not available were omitted from haplotype-based analyses.

Diversity and Population Structure

Observed (H_O) and expected (H_E) heterozygosity and deviation from Hardy–Weinberg equilibrium were calculated separately

for each population using the "genetics" package in R (R Core Team 2012). The inbreeding coefficient F_{IS} was calculated as $(H_E - H_O)/H_E$.

The prcomp function in R was used to perform PCA of SNP genotypes. The number of significant principal components was estimated based on the Tracy-Widom distribution (Patterson et al. 2006). Individuals were assigned to groups based on significant principal components (PCs) using Ward clustering via the hclust function of R (van Heerwaarden et al. 2010).

Pairwise F_{ST} (Weir and Cockerham 1984) was calculated for each pair of genetic groups using all polymorphic SNPs. The relationships between genetic, geodetic, and altitudinal distance were evaluated using Mantel tests and partial Mantel tests in the R package "vegan" (Oksanen et al. 2011). The significance of correlations between pairwise genetic distances measured as $F_{ST}/(1 - F_{ST})$ and matrices of both geodetic distance and altitude differences among populations were estimated using 9999 permutations of rows and columns of the distance matrices.

Admixture and population structure were estimated using the software package STRUCTURE (Falush et al. 2003) based on genotypes at 10,000 random SNPs. This subset is sufficient to represent population structure and was used to shorten the computational time required by STRUCTURE. Data from 279 inbred lines (Cook et al. 2012) were used in analyses that included cultivated maize. STRUCTURE was run under the admixture and correlated allele frequency model. Two independent runs of 2,500 burn-in MCMC iterations followed by 20,000 retained iterations were performed for each number of groups (K) from 2 to 6 for analyses including maize and from 2 to 9 for analyses including only teosinte samples. Based on comparisons of independent runs, 20,000 iterations were sufficient for convergence. Results were inspected with STRUCTURE HARVESTER v0.6.8 (Earl and vonHoldt 2012) and visualized using DISTRUCT (Rosenberg 2004).

In initial analyses the Ahuacatitlan population appeared to be intermediate between subspecies. Its phylogenetic placement relative to *mexicana* and *parviglumis* was evaluated based on the configuration of shared derived alleles using the D statistic of Green et al. (2010) under three different trees. D is a measure of deviation from the expected number of discordant gene trees given a suggested genealogy. For each chromosome, a haplotype was sampled from each of three groups (Ahuacatitlan, *parviglumis* excluding Ahuacatitlan, and *mexicana*), and the D value was calculated based on all sites that were polymorphic among the three groups. Sampling was repeated 1000 times to obtain the distribution of D. Homozygous SNPs from the sister genus *Tripsacum* were used to determine the ancestral state.

Haplotype Sharing and Linkage Disequilibrium

Haplotypes were inferred with fastPHASE (Scheet and Stephens 2006) using known haplotypes of 18 teosinte inbred lines (Chia et al. 2012) and default parameters. Sites with residual heterozygosity in the inbred lines were excluded. Segments of identity-by-state (IBS) and runs of homozygosity were identified using GERMLINE (Gusev et al. 2009), with a seed segment size of 50 SNPs and allowing zero heterozygous and homozygous mismatches.

The software TASSEL (Bradbury et al. 2007) was used to estimate linkage disequilibrium (LD) (r^2 and its P value) using phased data from all individuals and pairs of sites with minimum allele frequency >0.1. Blocks of LD were identified based on visual inspection of LD plots.

Candidate SNP Identification

Due to evidence of admixture, the Ahuacatitlan population was excluded from candidate SNP analyses. BAYENV (Coop et al. 2010) was used to evaluate the correlation between environmental variables and allele frequencies of individual SNPs. A random set of 10,000 SNPs was used to construct three covariance matrices: all populations and separately for each subspecies without the Ahuacatitlan *parviglumis* population. The maximum observed differences between two independent (50,000 iterations each) estimates of the covariance matrices were always less than approximately 10% of the smallest estimated covariance, indicating good convergence between runs.

Seventy-six environmental variables were analyzed, including 8 soil variables, 19 bioclimatic variables, monthly precipitation and monthly mean, maximum and minimum temperature, and altitude (supplementary table S2, Supplementary Material online). For climatic variables, 30 arc-second (~1 km) resolution climate data were downloaded from www.worldclim.org (last accessed August 12, 2013), and DIVA-GIS (Hijmans et al. 2001) was used to extract climate data for the population locations. Data for three key soil qualities (rooting conditions, oxygen availability to roots, and workability) that varied among populations were downloaded from the Harmonized World Soil Database, and data on five varying (cracking clays, volcanic, top soil clay, top soil sand, and top soil loam) key modifier layers (Sanchez et al. 2003) were downloaded from www.harvestchoice.org (last accessed August 12, 2013). All variables were standardized to a mean of 0 and standard deviation of 1. The dimensionality of environmental data was reduced using PCA (prcomp in R), and the PCs that captured 95% of the variance in environments among populations were used for analysis in BAYENV (supplementary table S3, Supplementary Material online).

Five independent BAYENV runs with 1,000,000 iterations were used to identify SNPs associated with these PCs. SNPs were considered candidates if they showed average Bayes factors across runs in the 99th percentile and were consistently in the 95th percentile of each run. A gene was considered a candidate when a candidate SNP was in the transcribed part of the gene.

To obtain the expected distribution of heterozygosity and the hierarchical differentiation statistics F_{CT} (among subspecies) and F_{ST} (among 20 populations, excluding Ahuacatitlan), 100,000 coalescent simulations were conducted under a hierarchical island model of two groups of 100 demes using the software Arlequin (Excoffier and Lischer 2010). The maximum expected heterozygosity for simulations was 0.5. F_{ST} and F_{CT} outliers were identified by comparing observed values to this model-based distribution of heterozygosity and differentiation.

The data consist of multiple populations and estimation of the overall deviation from random mating due to population structure (F_{ST}) may obscure the signal of differentiation in individual populations. To identify SNPs underlying differentiation at the level of individual populations, F statistics were calculated in a hierarchical framework using the R package "hierfstat" (Yang 1998; Goudet 2005). Variance components were calculated for three levels within each subspecies: population, focal, and individual. The focal component was calculated for each population and locus, yielding

$$F_{FT} = \frac{\hat{\sigma}_F^2}{\hat{\sigma}_F^2+\hat{\sigma}_S^2+\hat{\sigma}_I^2+\hat{\sigma}_E^2},$$

the proportion of total variance due to differentiation between a focal population and all other populations, where $\hat{\sigma}_F^2$, $\hat{\sigma}_S^2$, $\hat{\sigma}_I^2$, and $\hat{\sigma}_E^2$ are the estimates of variance between focal and other populations, among all populations, among all individuals, and error, respectively. Negative variance components were set to zero. For each population, SNPs with observed F_{FT} values above the 99th percentile were considered candidates.

As a complement to the above approaches focusing on individual SNPs, we additionally applied the pairwise haplotype sharing (PHS) method of Toomajian et al. (2006). The PHS statistic estimates the average genetic map length of shared haplotypes around a given allele, with significance evaluated by comparison to shared haplotype length for that position across all individuals. We used the same genetic map and phased haplotypes as in the analysis of runs of homozygosity (ROH) in GERMLINE. The PHS test was conducted first for all samples and then separately for each individual population.

Trait Association Analysis

To investigate the potential relationship between phenotypic traits and local adaptation, we asked whether putatively adaptive SNPs identified here showed enrichment for loci significantly associated with traits in a maize diversity panel. Association mapping tests were carried out on 278 inbred maize lines (supplementary text S1, Supplementary Material online) from the association panel described in Flint-Garcia et al. (2005). These lines were genotyped using the Illumina MaizeSNP50 array (Cook et al. 2012) and phenotyped for 36 traits (127 trait/environment combinations) (Hung et al. 2012). The 51,253 SNPs with minor allele frequency >0.05 were tested against at least one of the traits. Associations were tested using the model

$$\hat{G} = \mathbf{1}\mu+M\theta+S\beta+Zu+\varepsilon$$

where $\hat{G}$ is the estimated genetic value of a given trait, μ the trait mean, M the tested SNP, θ the SNP effect, S the matrix of the panel structure as estimated by Flint-Garcia et al. (2005), β the vector of structure effects, Z an incidence matrix, u a vector of random effects assumed to follow a distribution $N(0, \sigma_g^2 K)$,where K is the genetic variance-covariance matrix modeled by a shared allele matrix, and ε is the model residual assumed to follow $N(0, \sigma_\varepsilon^2 I_n)$. Only SNPs significantly associated to a given trait with P values <0.01 were retained for further analyses. To test whether enrichment for flowering time loci was affected by a loss of power due to correction for population structure, we also conducted association analyses for flowering time traits using a simplified model that did not include population structure and coancestry.

Enrichment Analyses

Enrichment of candidates in genic (within the transcribed region of a gene) versus nongenic regions and nonsynonymous versus synonymous classes was calculated by sampling the number of SNPs in each candidate list randomly from the genome 1000 times. Enrichment was assessed both with and without SNPs in the four inversion polymorphisms (*Inv1n* [Fang et al. 2012], *Inv4m*, *Inv9e*, and *Inv9d*) that were identified based on LD patterns (see Results for details). Enrichment of SNPs within inversions was further inspected by ranking SNPs in each candidate list and calculating the proportion of SNPs within inversions in each candidate list. The joint effect of all inversions across all candidate lists was determined by assigning each SNP a maximum rank (more significant ranks being higher) across all candidate lists and calculating the proportion of SNPs from inversions above the 99th quantile. Bootstrapping was used again to determine statistical significance.

Results

Genomic Diversity in Teosinte

We sampled 21 populations of the wild subspecies *parviglumis* and *mexicana* from across their native ranges in central and southern Mexico (fig. 1). Ten to 12 individuals from each population were genotyped using the Illumina MaizeSNP50 chip. After quality control, the full data set consisted of 248 samples genotyped at 36,719 SNPs. Although SNPs on the MaizeSNP50 chip were ascertained in a small sample of maize lines, ascertainment appears to have had minimal effect on analyses performed here, as shown by comparisons of minor allele frequencies between subspecies (supplementary fig. S1, Supplementary Material online) or between SNPs with

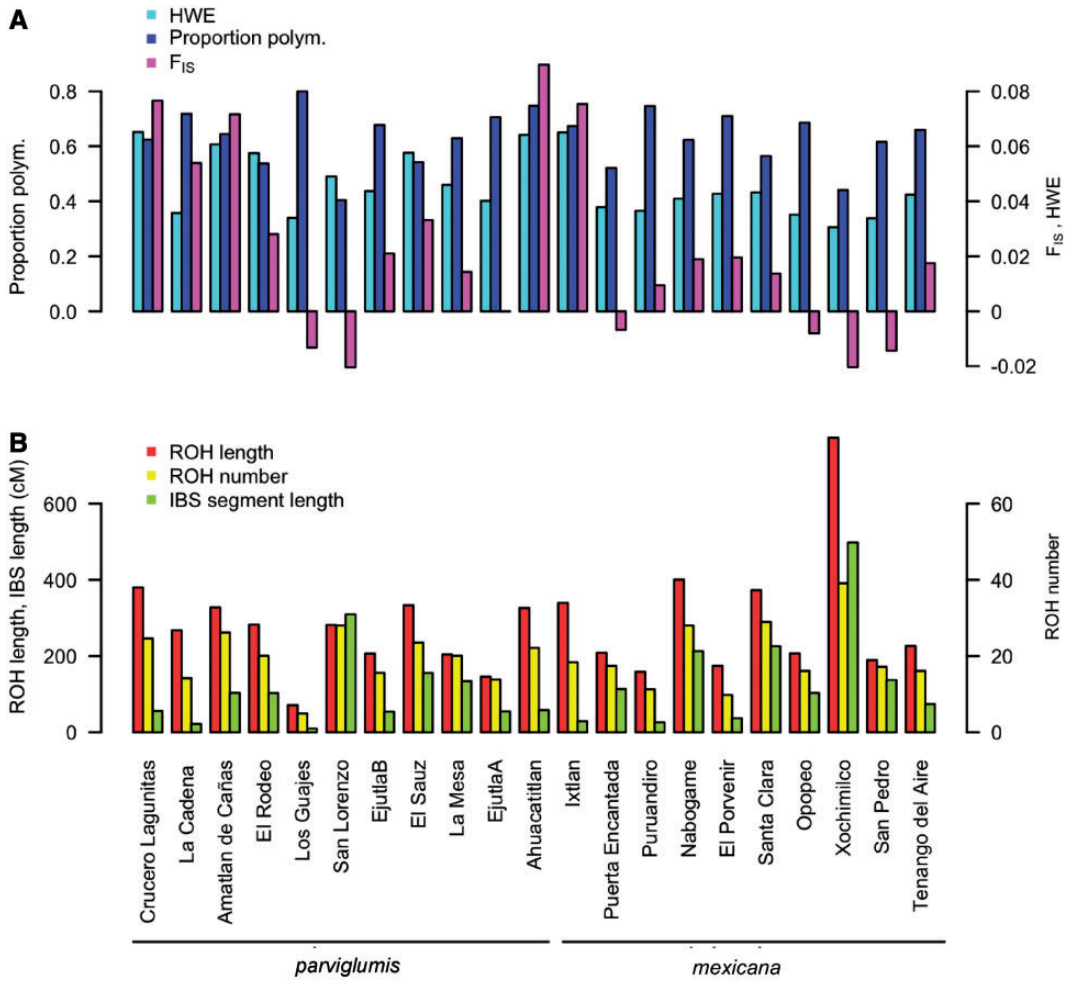

Fig. 2.—Diversity statistics. (*A*) Proportion of SNPs deviating from Hardy–Weinberg Equilibrium (HWE), proportion of polymorphic SNPs, and mean inbreeding coefficient F_{IS}. (*B*) Length and number of ROH and average pairwise length of genomic segments IBS.

different-sized ascertainment panels (supplementary fig. S2, Supplementary Material online).

Overall, heterozygosity differed only slightly between the two subspecies, but considerable variation was evident among populations in both taxa (fig. 2 and supplementary table S4, Supplementary Material online). Both *parviglumis* and *mexicana* showed generally high levels of heterozygosity and little evidence of inbreeding. Within *parviglumis*, the population of Los Guajes showed high diversity, an inbreeding coefficient (F_{IS}) close to zero, no deviation from Hardy–Weinberg equilibrium, and relatively short runs of ROH (supplementary fig. S3, Supplementary Material online), as might be expected of a large, outcrossing population. Other populations, however, showed evidence of recent demographic changes: diversity in San Lorenzo *parviglumis* was only half that of Los Guajes, and *mexicana* individuals from Xochimilco were characterized by extremely long ROH and extensive within-population haplotype sharing, consistent with a recent population bottleneck.

Patterns of LD varied greatly along the genome. Although LD generally decays quickly in teosinte (Hufford et al. 2012) and the median LD within 5-kb windows in our data is low ($r^2 = 0.04$), we observed several discrete blocks of elevated LD in multiple populations. We identified four large (>10 Mb) regions of high ($r^2 \geq 0.2$) LD (fig. 3 and supplementary fig. S4, Supplementary Material online). Three of these regions appear to correspond to inversions previously described cytologically (*Inv9e*; Ting 1964), in mapping populations (*Inv4m,* Mano et al. 2012), or by population genetic analysis (*Inv1n*, Fang et al. 2012; Hufford et al. 2012). The size of these blocks and their chromosomal location effectively rules out recent selective sweeps or centromeric regions as alternative explanations, and we interpret all four as inversion polymorphisms. Clear haplotype structure was observed in IBS analysis for three of these putative inversions, and simple genetic-distance-based clustering, including *Tripsacum* and maize, suggested that the nonmaize haplotype was likely the derived state (supplementary fig. S5, Supplementary Material

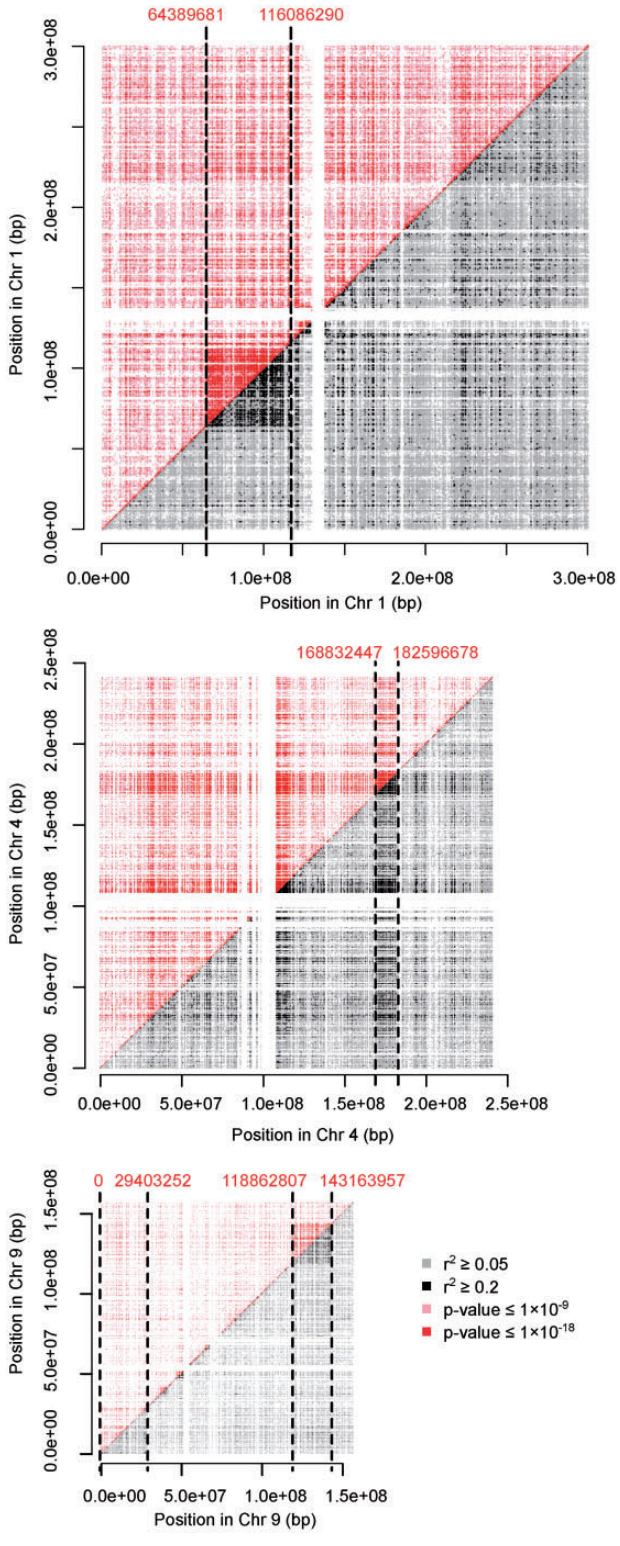

Fig. 3.—LD reveals structural rearrangements in teosinte. Shown are LD (r^2, red) and permutation *P* value (black) for pairs of SNPs across chromosomes 1, 4, and 9. Dashed black lines delineate the likely boundaries of structural variants discussed in the text. *P* value and r^2 cutoffs were chosen to visualize the LD blocks against the background level of LD.

online). The inverted and noninverted haplotypes of *Inv9d* were less clearly demarcated, however, due to low frequency and high diversity of the putative inverted haplotype (supplementary fig. S5, Supplementary Material online).

Population Structure

Allele frequency in teosinte varied at multiple hierarchical levels. PCA of our SNP data revealed the strongest signal of population structure between the two subspecies (PC1, ~12% of the variation explained; fig. 4*A*). Differentiation between subspecies was also evident in results from STRUCTURE

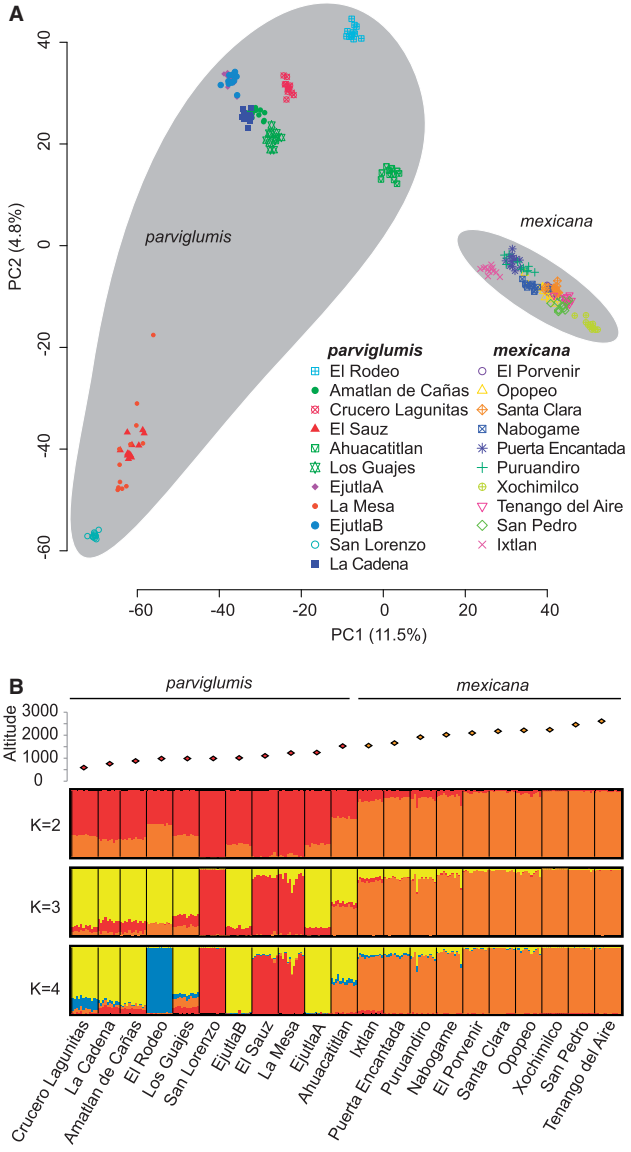

Fig. 4.—Population structure in teosinte. (*A*) Principal component analysis of all individuals, labeled according to the sampled population. (*B*) STRUCTURE results for all individuals. Individuals are grouped by population and populations ordered by increasing altitude.

(fig. 4*B*), haplotype sharing (supplementary figs. S6 and S7, Supplementary Material online), and higher pairwise F_{ST} between subspecies (average of all *parviglumis–mexicana* pairs: 0.33) than within subspecies (among *parviglumis* populations 0.24; among *mexicana* populations 0.23) comparisons. Putative inversion polymorphisms also separated the subspecies: *Inv9d* and *Inv9e* were found only in *mexicana*, whereas the derived haplotype at *Inv4m* showed a strong frequency difference between subspecies (supplementary fig. S8, Supplementary Material online).

Additional levels of structure were observed within subspecies as well (supplementary fig. S9, Supplementary Material online) with STRUCTURE and PCA analysis identifying groups of related populations largely consistent with previous work (Fukunaga et al. 2005). In all, the 20 significant principle components identified 21 clusters that generally corresponded to sampling locales (fig. 4*A*). Three populations of *parviglumis* did not follow this trend however: the two Ejutla populations formed a single genetic cluster and the Ahuacatitlan population split into two clusters.

Both geodetic distance (Mantel test *P* value: 0.006, Mantel statistic $r = 0.40$) and altitude (*P* value: 0.0008, $r = 0.37$) appeared to correlate with population differentiation in the combined data set (supplementary fig. S10*A*, Supplementary Material online; partial Mantel tests for geodetic distance *P* value: 0.0005, $r = 0.39$; altitude *P* value: 0.004, $r = 0.41$), consistent with analyses suggesting an important role for altitude in patterning diversity in a range-wide analysis of teosinte (Bradburd et al. 2013). Differentiation within each subspecies significantly correlated with geodetic distance (*P* value: 0.004, $r = 0.55$ in *parviglumis*; *P* value: 0.046, $r = 0.48$ in *mexicana*; supplementary fig. S10*B* and *C*, Supplementary Material online). However, this trend was driven mostly by the genetically and geographically distant populations of Nabogame and El Rodeo and was no longer significant after their removal.

Finally, both STRUCTURE and PCA results suggested that the Ahuacatitlan population of *parviglumis* was admixed (fig. 4). Consistent with this, analysis of haplotype sharing showed that the Ahuacatitlan population shared fewer long haplotypes with *parviglumis* and more long haplotypes with *mexicana* than any other *parviglumis* population (supplementary figs. S6 and S7, Supplementary Material online). The average length of shared haplotypes between Ahuacatitlan and *mexicana* (mean 4.8 cM) was, however, shorter than the average length of haplotypes shared among *mexicana* populations (mean 30.7 cM), arguing against extensive recent admixture. To explicitly test the origin of Ahuacatitlan, we applied a test of the configuration of shared derived alleles (Green et al. 2010) with populations of each of the two subspecies. The observed distribution of the D test statistic was inconsistent with models in which Ahuacatitlan is ancestral to both *parviglumis* and *mexicana* or a sister group to *mexicana* ($P < 0.05$ for both; supplementary fig. S11, Supplementary Material online), whereas the model in which Ahuacatitlan was sister to *parviglumis* could not be rejected (*P* value: 0.2).

Candidate SNP Identification

We applied four approaches to identify SNPs underlying local adaptation; candidate SNPs and genes from each of these approaches are listed in supplementary table S5, Supplementary Material online. We first applied the environmental association method implemented in BAYENV (Coop et al. 2010) to principal components summarizing 95% of the variation found among populations from a set of 76 climate and soil variables. Bayes factors estimated for a joint

Table 1
Summary of Environmental Correlation and Differentiation Outlier Results

Analysis	Variable	Major Loadings	BF, 99th[a]	No. Candidate SNPs	No. Candidate Genes
Both	PC1	Altitude, temperature	116.8	262	162
	PC2	Temperature seasonality, soil quality and precipitation	400.0	359	222
	PC3	Precipitation	40.4	308	201
	PC4	Topsoil variables and precipitation seasonality	39.4	229	145
	PC5	Mean diurnal range of temperature plus some soil variables	72.9	291	184
	PC6	Mean diurnal range of temperature plus some soil variables	10.2	225	151
parviglumis	PC1	Altitude, temperature	2.2	60	36
	PC2	Temperature range and seasonality	6.4	173	118
	PC3	Soil type and precipitation	2.0	102	69
	PC4	Monthly precipitation and mean diurnal range of temperature	4.8	42	24
mexicana	PC1	Altitude, temperature	6.9	108	69
	PC2	Temperature seasonality and range and precipitation	52.3	309	204
	PC3	Precipitation and volcanic soil	31.1	27	16
	PC4	Top soil variables and temperature variability	3.8	93	58
F_{CT}				731	728
F_{ST}				1363	411

[a]Bayes factor (BF) at 99th percentile of distribution.

data set of *parviglumis* and *mexicana* identified 1,598 SNPs associated with one or more principle component(s), while 370 SNPs and 533 SNPs were identified for each subspecies respectively (table 1 and supplementary figs. S12–S14, Supplementary Material online).

Our second approach identified locally adapted SNPs using an F_{ST} outlier method (Lewontin and Krakauer 1973; Excoffier et al. 2009) to scan for SNPs showing an excess of differentiation between subspecies (F_{CT}) or among populations (F_{ST}). Based on simulations under the hierarchical model, 731 F_{CT} (2.0%) and 1363 F_{ST} (3.7%) outliers were identified at a *P* value cutoff of 0.01. A striking peak in differentiation between both subspecies and populations was observed within inversion *Inv4m*, which showed 12-fold and 6-fold enrichment of outlier SNPs for F_{CT} and F_{ST}, respectively. Both sets of outliers (fig. 5 and supplementary table S5, Supplementary Material online) differed substantially from SNPs identified by BAYENV, and correlations between Bayes factors and differentiation were low (supplementary table S6, Supplementary Material online).

Third, we estimated the proportion of variation due to differentiation in a single focal population, F_{FT} (see Materials and Methods). Mean F_{FT} across populations was lower than 0.1, suggesting modest deviation from random mating due to individual populations. SNPs in the extreme 1% tail of the F_{FT} distribution generally showed little overlap or correlation with F_{ST} or F_{CT} values, but populations at range extremes such as El Rodeo and Nabogame had higher F_{FT}, and high F_{FT} SNPs from these populations were also overrepresented among F_{ST} (Wilcoxon rank-sum test *P* values <0.001) and BAYENV outliers (Wilcoxon rank-sum test *P* values <0.001 for PC1, PC2, and PC3). This effect was most evident in Nabogame, where F_{FT} outliers constituted approximately 9% of all PC2 outliers for BAYENV.

Finally, to complement the above approaches focusing on individual SNP frequencies, we implemented the PHS test (Toomajian et al. 2006) to identify regions with extensive haplotype sharing as might be expected from a recent partial sweep. When the PHS test was applied to our entire sample (supplementary fig. S15, Supplementary Material online),

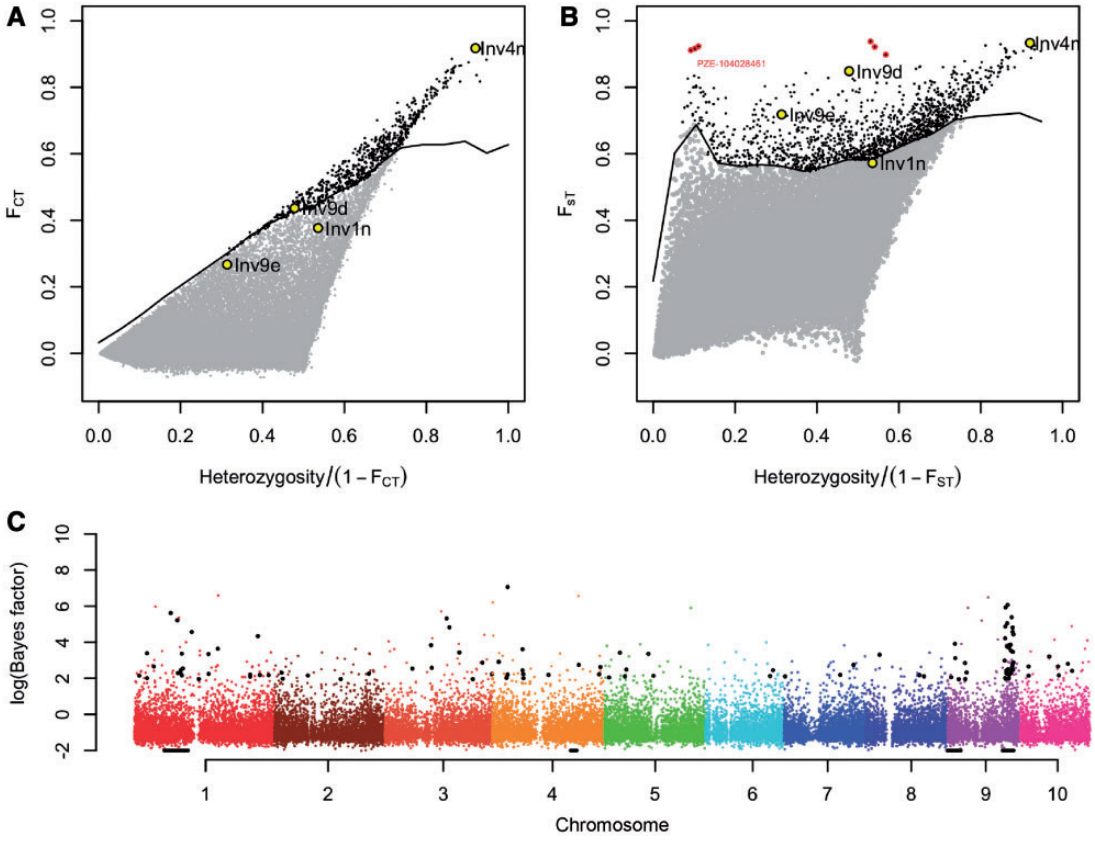

Fig. 5.—Differentiation and environmental correlation in inversions compared with genome-wide distribution of test statistics. Joint distribution of F_{CT} (*A*) and F_{ST} (*B*) versus heterozygosity under a hierarchical island model for all SNPs. The black line indicates the 1% tail based on simulations. Values for each inversion when treated as a single locus indicated with yellow. Extreme outlier loci indicated with red. (*C*) Bayes factors for PC1 in *mexicana*. Values are plotted across all 10 chromosomes, with each chromosome in a different color. Black dots represent outlier SNPs, and black horizontal bars below the plots indicate the positions of inversions.

outliers (P value cutoff 0.01) showed significant overlap (P value < 0.001) with PC1 BAYENV, F_{FT}, and F_{ST} outliers, due primarily to haplotype structure created by the putative inversions on chromosomes 4 (position 168,832,447–182,596,678) and 9 (positions 0–29,403,252 and 118,862,807–143,163,957). Significant overlap was also observed between F_{FT} outliers and PHS outliers identified in tests of individual populations (11 out of 21 populations had P values <0.01 based on permutations). These combined results indicate that SNPs identified as differentiated in a specific population often reside in a region with unexpectedly long shared haplotypes.

Characterization of Candidate Loci

We assessed additional evidence for local adaptation at candidate loci by further dissecting signals in putative inversions and gauging enrichment of candidates in described functions. All four putative inversions contained an excess of differentiated SNPs or SNPs associated with environmental variables (fig. 6 and table 2). To partially control for the nonindependence of loci within inversions, each SNP was given a maximum rank across all environmental association candidate lists (see Materials and Methods for details). Inversions were 2-fold enriched for candidate SNPs (P value < 0.001), containing 5.6% of all SNPs but 11% of SNPs in the 99th percentile of the maximum rank distribution of these lists.

To further control for nonindependence of SNPs within putative inversions, we conducted an additional BAYENV analysis with inversions treated as single biallelic loci. We assigned each individual a genotype based on neighbor-joining trees of SNPs within the inversion (supplementary fig. S5, Supplementary Material online). All four putative inversions were identified in environmental association analysis when treated as biallelic loci. Three of these inversions (*Inv1n*, *Inv4m*, and *Inv9d*) showed altitudinal clines (supplementary fig. S8, Supplementary Material online), though the signal at *Inv1n* was less clear than originally reported in Fang et al. (2012) as the derived (inverted) arrangement was absent from the low elevation Crucero Lagunitas population of *parviglumis*.

We tested candidate SNPs for enrichment in functional categories (genic vs. nongenic, synonymous vs. nonsynonymous) as well as for association with maize phenotypic traits (Flint-Garcia et al. 2005; Hung et al. 2012), putative adaptive introgressions from *mexicana* into maize (Hufford et al. 2013), and ±5 Mb around centromeres (Wolfgruber et al. 2009). PHS outliers were the only set of candidates enriched for genic SNPs (P value <0.001 for outliers based on the total sample, with 11 of 21 populations having P <0.01). In contrast, F_{ST} outliers were strongly enriched for nongenic SNPs (P values < 0.001, supplementary table S7, Supplementary Material online), and both F_{ST} and F_{CT} outliers were farther from genes than other SNPs (P values < 0.001, two-sample Kolmogorov–Smirnov test). F_{ST} outliers were also enriched for SNPs associated with the architecture of the male inflorescence in a diverse maize panel (Flint-Garcia et al. 2005;

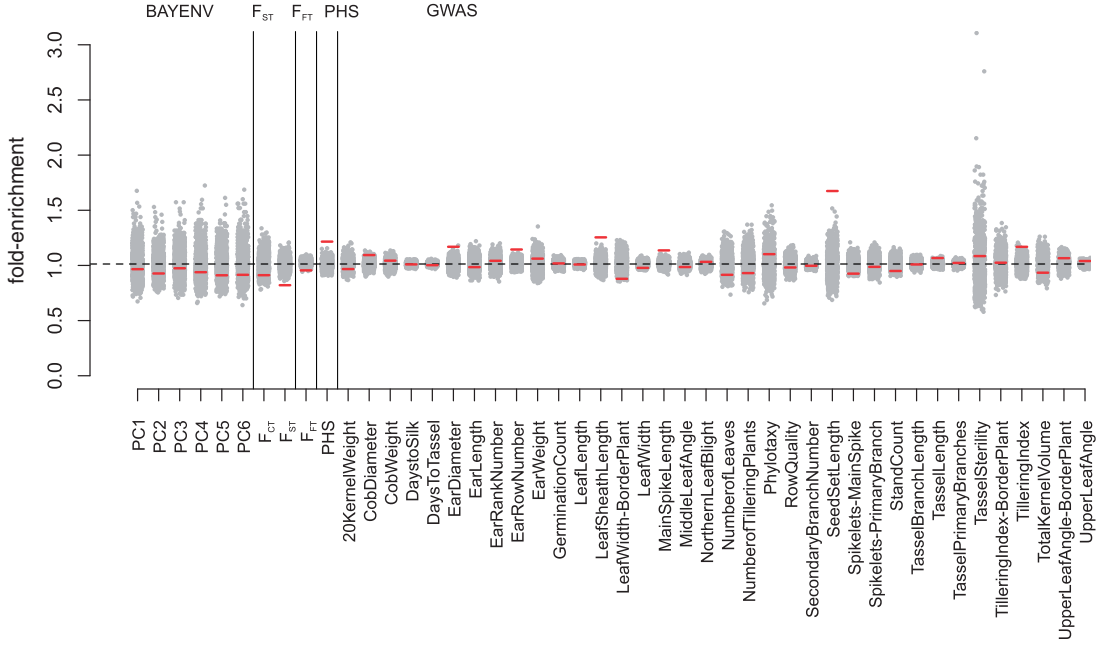

Fig. 6.—Fold enrichment of observed ratios of genic/nongenic SNPs among candidates. Fold of enrichment for each set of candidates (red line) from BAYENV, F_{ST}, and F_{CT}, SNPs significant in any of the population-based F_{FT} and PHS analysis, and phenotypic association analysis (GWAS). Expected distribution of fold of enrichment ratios (gray dots) was obtained by random sampling of the same amount of SNPs as in the candidate set 1000 times and calculating the ratio and fold enrichment. SNPs within inversions were excluded.

Table 2
Enrichment of Sets of Candidate SNPs within Four Inversions

Analysis	Variable	*P* Value[a]					Biallelic
		Inv1n	*Inv4m*	*Inv9d*	*Inv9e*	All Inversions	
Both	PC1	<0.001 (4.5)	0.003 (3.7)	<0.001 (7.6)	0.027 (2.2)	0.0001 (3.3.)	*Inv9d*
	PC2	0.032 (1.7)	0.824	<0.001 (3.9)	0.331	0.0016 (1.8)	
	PC3	<0.001 (5.7)	1.000	0.819	0.186	0.0001 (2.2)	*Inv1n*
	PC4	0.002 (2.7)	0.574	0.001 (3.1)	<0.001 (4.0)	0.0001 (2.5)	*Inv9e*
	PC5	0.440	1.000	1.000	0.938	0.9942	
	PC6	0.672	1.000	0.859	0.687	0.9745	
parviglumis	PC1	0.450	1.000	0.290	1.000	0.0181 (1.5)	
	PC2	0.616	0.016 (3.2)	0.045 (2.3)	0.728	0.1555	
	PC3	0.261	1.000	1.000	0.812	0.3764	
	PC4	0.271	1.000	1.000	0.474	0.4543	
mexicana	PC1	0.001 (3.8)	1.000	<0.001 (33.0)	0.095	0.0001 (3.2)	*Inv9d*
	PC2	0.748	0.694	<0.001 (4.3)	0.027 (2.0)	0.0007 (1.8)	
	PC3	0.136	1.000	1.000	1.000	7.99	
	PC4	0.002 (3.9)	1.000	1.000	<0.001 (8.2)	0.0001 (2.0)	*Inv9e*
F_{CT}		<0.001 (2.9)	<0.001 (11.7)	0.049 (1.6)	0.705	0.0001 (3.4)	*Inv4m*
F_{ST}		<0.001 (2.7)	<0.001 (6.3)	<0.001 (1.8)	0.866	0.0001 (3.2)	*Inv4m*

Note.—Inversions significant when analyzed as a single biallelic locus are listed in the column "biallelic."
[a]For each *P* value <0.05, the magnitude of enrichment is reported in parenthesis. *P* values are based on bootstrapping.

Cook et al. 2012; supplementary table S8, Supplementary Material online). SNPs significantly associated with flowering time, ear size, and stand count (plant survival and density) were overrepresented among both F_{CT} and F_{ST} outliers (supplementary table S9, Supplementary Material online), but did not show enrichment in an association model that took into account structure and relatedness (both K + Q; see Materials and Methods). Among genes that showed evidence of introgression from *mexicana* into maize, both F_{ST} (*P* value < 0.001) and F_{CT} (*P* value < 0.001) candidates as well as SNPs associated with PC2 (temperature range) within *mexicana* (*P* value 0.0061) were more common than expected by chance (Hufford et al. 2013). Excluding SNPs within inversions did not qualitatively change any of these results (data not shown).

In centromeres, weak but significant enrichment of environmentally correlated SNPs was observed for PC1 (2.7-fold, $P < 0.001$) and PC2 (1.5-fold, $P = 0.046$) in joint analysis, and PC3 in *mexicana* (4.3-fold, $P = 0.009$) (supplementary figs. S12–S14, Supplementary Material online). Both F_{ST} and F_{CT} outliers were 5-fold enriched in centromeres ($P < 0.001$), mostly due to centromeres on chromosomes 2 and 4, where enrichment was 21- and 17-fold for F_{ST} and F_{CT}, respectively (supplementary fig. S17, Supplementary Material online).

Discussion

Complex Population Structure

The presence of population structure is a crucial consideration when determining the potential for local adaptation. Population structure can bias estimates of demographic history and the inference of selection through its effect on the allele frequency spectrum (Siol et al. 2010). In addition, Eckert et al. (2010b) point out that when environmental gradients and population structure coincide, environmental correlation methods may suffer from increased rates of false positives and negatives if the analysis under- or overcorrects for population structure. Previous theoretical work under simple island (Städler et al. 2009) or stepping stone (De and Durrett 2007) models has shown that biases in demographic inference can be ameliorated by sampling single individuals from each of many populations. However, such a sampling scheme does not take into account uneven or hierarchical population structure (St. Onge et al. 2012).

Our results indicate that population structure in teosinte is complex and affected by multiple factors. Across all populations, the effects of altitude on population structure are significant, even after correcting for geodetic distance (supplementary fig. S10A, Supplementary Material online). The *mexicana* populations of Santa Clara and Opopeo, for example, show higher F_{ST} with the La Cadena *parviglumis* population 60 km away than with distant *parviglumis* populations at the same elevation (supplementary fig. S16, Supplementary Material online). These results are in agreement with other studies that have noted the important role of altitude in determining genetic distance, genome size, and morphological characters in *Z. mays* subspecies (Smith et al. 1981; Poggio et al. 1998; Buckler et al. 2006; van Heerwaarden et al. 2011; Bradburd et al 2013). There is also clear hierarchical population structure (fig. 4 and

supplementary fig. S9, Supplementary Material online), with divisions between subspecies and among previously identified groups within subspecies (Fukunaga et al. 2005). Within subspecies, the effect of altitude is no longer statistically significant, but isolation by distance continues to explain a meaningful portion of the genetic structure observed in both taxa (supplementary fig. S10*B* and *C*, Supplementary Material online). In addition to distance and altitude, stochastic founding events have likely played a role in patterning diversity. This is most evident in the extensive haplotype sharing within the geographically dispersed La Mesa/El Sauz/San Lorenzo group (supplementary fig. S6, Supplementary Material online). The importance of founder events is consistent with previous analyses by Fukunaga et al. (2005) and Buckler et al. (2006) who observed that genetic distance was more correlated with specific dispersal routes than with geodetic distance.

Population structure had a considerable effect on many of our outlier detection approaches. *Parviglumis* and *mexicana* do not conform to a simple two-level hierarchical island model where gene flow between populations within subspecies would be equal. The impact of this structure when determining candidates for local adaptation was most clearly seen in a subset of strongly differentiated populations. F_{FT} outlier SNPs specific to the isolated El Rodeo and Nabogame populations have smaller F_{ST} *P* values than F_{FT} outliers from other populations (Wilcoxon rank-sum test, $P < 0.001$). Similarly, the Nabogame *mexicana* population has extreme temperature and precipitation, and SNPs that are strongly differentiated in Nabogame (F_{FT} outliers) have significantly higher Bayes factors for PC2 (temperature seasonality) in the BAYENV analysis than F_{FT} outliers from other *mexicana* populations (Wilcoxon rank-sum test, $P < 0.001$). While these findings may reflect true local adaptation, they should be interpreted with caution since such geographically restricted environmental variation has been shown to cause false positives in association studies even when population structure is taken into account (Mathieson and McVean 2012). This may be one explanation for the observation that new mutations with narrow geographic distributions were enriched among climate-associated loci in *Arabidopsis* (Hancock et al. 2011b). Because most plants likely show complex patterns of structure (Loveless and Hamrick 1984), it is clear that careful population level sampling and consideration of population structure will continue to be important in studies of plant adaptation and evolutionary history.

In addition to providing insight regarding local adaptation, our examination of population structure identified a putative hybrid zone between *parviglumis* and *mexicana*. While hybridization between these teosinte taxa is known to occur (Fukunaga et al. 2005; van Heerwaarden et al. 2011), details of admixture within individual populations have not been well documented. Our data suggest that our highest elevation *parviglumis* population, Ahuacatitlan, is extensively admixed, with mean assignment probabilities to *mexicana* of approximately 50% across individuals in the population. Patterns of haplotype sharing (supplementary figs. S6 and S7, Supplementary Material online) and derived allele counts (supplementary fig. S11, Supplementary Material online), however, support a model of continual gene flow with *parviglumis* over a relatively long period of time. Several of the *parviglumis* individuals sampled by Ross-Ibarra et al. (2009) are from localities very near to Ahuacatitlan, likely contributing to these authors' inference of continuous gene flow between *mexicana* and *parviglumis*. Reports of hybrids from localities near Ahuacatitlan (Fukunaga et al. 2005), together with our results showing little admixture in other *parviglumis* populations in geographic proximity to *mexicana*, indicate the presence of a geographically restricted hybrid zone between *mexicana* and *parviglumis* occurring at mid elevations in the Eastern Balsas region.

Adaptive Inversion Polymorphisms

Cytological studies of both *mexicana* and *parviglumis* have previously identified a number of inversion polymorphisms (Ting 1964; Wilkes 1967; Kato 1975). In our genotyping data, we observe four blocks of LD that we interpret as large inversions. Together they comprise approximately 5% of the maize reference genome, and their role in patterning both molecular and phenotypic variation (Fang et al. 2012) among populations is significant.

Three of the inversion polymorphisms we have identified here based on LD have previously been observed. Inversion *Inv1n* was first identified by Hufford et al. (2012) and its population genetics subsequently described by Fang et al. (2012). *Inv9e* was identified cytologically by Ting (1964), who found the inverted arrangement in both of the *mexicana* populations in which we see the derived haplotype. Finally, the LD block we identify as *Inv4m* is at or near the physical positions of several inversions identified cytologically in maize and teosinte (Ting 1964; Wilkes 1967) and by marker map order in *Zea nicaraguensis* (Mano et al. 2012). Synteny maps identify the proximal breakpoint of *Inv4m* as an ancient chromosomal fusion involving the homeolog of the distal end of the telomere of chromosome 5 (Schnable et al. 2009; Wang and Bennetzen 2012), and the breakpoint appears common to several distinct inversions in maize (Doyle 1994). In addition, there appears to be long-distance LD between *Inv4m* and the centromere of chromosome 4 (fig. 3). All in all, this suggests that the region of *Inv4m* may be prone to structural rearrangements.

All four observed inversions appear to play a role in local adaptation: they show significant enrichment for SNPs with high Bayes factors for PC1 (table 2) and are characterized by altitudinal clines in haplotype frequency (supplementary fig. S8, Supplementary Material online) reflecting environmental differences such as temperature. Inversion *Inv1n* is enriched

for SNPs with high Bayes factors for PC1, PC3, and PC4, and the clinal patterns observed are broadly consistent with those of Fang et al. (2012). Extreme patterns of precipitation in Crucero Lagunitas and shared ancestry with the distant Oaxacan population of El Rodeo suggest that distinct selection pressures or unusual history may explain the surprising lack of *Inv1n* in this low-elevation population. *Inv4m* is associated with a striking peak in differentiation among subspecies (supplementary fig. S17, Supplementary Material online) and is also an extreme F_{ST} and F_{CT} outlier when treated as a single locus (fig. 5). This suggests an adaptive role for *Inv4m* at higher altitudes, consistent with evidence of introgression from *mexicana* into sympatric maize populations at *Inv4m* and its rarity in maize outside of the Mexican highlands (Hufford et al. 2013). Interestingly, quantitative trait loci (QTL) for differences between *parviglumis* and *mexicana* for pigment intensity and macrohair count—traits thought to be adaptive at high elevations—co-localize to this region as well (Lauter et al. 2004). *Inv9d* was only polymorphic in *mexicana*, but showed a strong enrichment of SNPs associated with PC1 (fig. 5*C* and supplementary fig. S14*A*, Supplementary Material online). Finally, inversion *Inv9e* showed 8-fold enrichment of SNPs associating with PC4 (top soil) within *mexicana* (table 2 and supplementary fig. S14*D*, Supplementary Material online). While all four inversions show evidence of natural selection, further experiments will be required to identify the precise targets of selection within the inversions and directly measure their effects on traits and fitness.

The number of polymorphic inversions observed here underscores the potential importance of inversions in plant local adaptation (Kirkpatrick 2010). The accumulation of locally adaptive loci in inversions has been predicted by theory (Kirkpatrick and Barton 2006), and is consistent with observations in a number of plant and animal taxa (Dobzhansky 1970; Etges and Levitan 2004; Huynh et al. 2010; Lowry and Willis 2010; Cheng et al. 2012). Although, in principle the observation of high differentiation in inversions could be due to the effects of background selection in regions of low recombination (Charlesworth et al. 1995), SNP correlations with environmental variables and clinal patterns of inversion haplotypes are difficult to explain under a model of background selection.

Although similar patterns of enrichment (excess of BAYENV, F_{ST}, and F_{CT} outliers) were observed in some centromere regions, the low levels of recombination observed in maize centromeres (Gore et al. 2009) imply that these results, as well as previous observations of haplotype differentiation (Hufford et al. 2012), may be due to the effects of selection on linked sites (Cutter and Payseur 2013) rather than on centromeres themselves.

Genome-Wide Patterns of Local Adaptation

Candidate SNPs identified here as outliers for environmental association or allele frequency differentiation show a pronounced enrichment in nongenic regions (fig. 6). This result stands in contrast to the genic enrichment found in similar analyses of both *Arabidopsis* (Hancock et al. 2011a) and human (Hancock et al. 2011b) data. One potential explanation for this difference is that the complexity of maize and teosinte genomes, which are >85% noncoding sequence (Schnable et al. 2009), may provide greater opportunity for the evolution of functional noncoding elements. Indeed, the causal polymorphisms underlying traits including flowering time (Salvi et al. 2007) and branching (Studer et al. 2011) have recently been identified as repeat insertions which impact gene regulation. The functional role of our candidate SNPs is supported by their enrichment among relevant phenotypic traits (supplementary tables S8 and S9, Supplementary Material online). For example, F_{CT} candidates are enriched for SNPs significantly associated with tassel morphology in maize, a major phenotypic difference between *parviglumis* and *mexicana* (Iltis and Doebley 1980). Both F_{CT} and F_{ST} candidates are enriched for SNPs significantly associated with flowering time, a common form of local adaptation observed among plant populations (Olsson and Ågren 2002; Lowry et al. 2008). We see little evidence that the observed nongenic enrichment is due to biases in the genotyping platform, as differences in heterozygosity between genic (0.275) and nongenic (0.281) SNPs are minor, and both our PHS outliers and some association analyses show enrichment for genic SNPs. These results do not argue that all noncoding DNA is functional; however, a recent genome-wide sampling of noncoding variants finds significantly fewer phenotypic associations than expected (Chia et al. 2012).

Although our candidate loci are not enriched for genes, we nonetheless identify a number of compelling genes showing signals of local adaptation (supplementary table S5, Supplementary Material online). A SNP in the gene *b1*, for example, correlated with PC1—largely made up of temperature and altitude—among *mexicana* populations. *b1* is a gene in the anthocyanin synthesis pathway that has been identified as a QTL for sheath color differences among *mexicana* and *parviglumis* (Lauter et al. 2004). Pigmentation has been suggested to be an important adaptive trait as a response to lower temperatures and changing ultraviolet light conditions (Galinat 1967; Barthakur 1974; Chalker-Scott 1999). Two SNPs in the 3′-UTR of the well-known domestication gene *teosinte branched1*, a locus known to determine plant architecture in *parviglumis* populations (Weber et al. 2007, 2008), show unusually strong patterns of differentiation and association with PCs related to temperature range and soil types (supplementary table S5, Supplementary Material online). Other candidate genes with a well-known function in maize include *abph1* that controls leaf arrangement (Jackson and Hake 1999) and *sh1* that encodes the production of sucrose synthase (Sheldon et al. 1983). Additionally, a synonymous SNP (PZE-104028461) in the

maize filtered gene GRMZM2G000471 was both an extreme outlier for F_{ST} and associated with PC2 in *mexicana*. The gene containing this SNP is orthologous to the *Arabidopsis* gene, At4g10380, *NIP5;1*, which encodes a boron channel (Kato et al. 2009) and has been identified as a candidate for adaptation to different soil types in *Arabidopsis lyrata* (Turner et al. 2008).

Finally, environmental association and frequency-based approaches produced sets of candidates with little overlap (supplementary table S6, Supplementary Material online, the overlap between PHS and environmental association was always less than 15 SNPs). Rather than indicating a lack of evidence for local adaptation, we suspect that the weak correspondence between methods is likely reflective of biological factors. For example, SNPs showing environmental association are only weakly correlated with F_{ST} and F_{CT} *P* values (supplementary table S3, Supplementary Material online), a result also reported in trees (Eckert et al. 2010a; Keller et al. 2012). Because F_{ST} outlier methods detect excess divergence among populations regardless of the distribution of environmental variation, they are likely more powerful for identification of loci related to factors that do not correlate with environmental variables included here. All of our measured environmental variables are abiotic, but biotic factors (e.g., herbivory, competition) undoubtedly play an important selective role and may not be perfectly reflected in our environmental PCs. For example, the presence of the *Zea* specialist leafhopper *Dalbulus maidis* is thought to depend on the local abundance of maize and proximity to bodies of water (Medina et al. 2012), and Moeller and Tiffin (2008) hypothesize localized selection pressure in herbivory as an explanation for evidence of local adaptation in their sequence analysis of immunity genes in *parviglumis.* Biotic interactions are likely critical in local adaptation in a number of species. For example, local pathogen diversity may play a larger role in human adaptation than climate (Fumagalli et al. 2011). Consistent with the importance of localized selection pressures, we note that many of our F_{FT} outliers, which show less overlap with other methods, nonetheless exhibit patterns of haplotype sharing suggestive of recent positive selection. A full characterization of local environments is, in most instances, unfeasible. Due to this limitation, comprehensive studies of local adaptation should combine methods that assess allele frequency association with measured variables with those that describe extreme population differentiation based on genetic data alone.

Supplementary Material

Supplementary text S1, tables S1–S9 and figures S1–S17 are available at *Genome Biology and Evolution* online (http://www.gbe.oxfordjournals.org/).

Acknowledgments

The authors thank Jeff Glaubitz for SNP annotations and mapping; Lauren Sagara for technical assistance with genotyping; Shohei Takuno for PHS perl scripts; and Norm Ellstrand, Pesach Lubinsky, and Mark Millard (USDA-ARS GRIN) for seeds. Yaniv Brandvain, Graham Coop, John Doebley, Andrew Eckert, Torsten Günther, Joost van Heerwaarden, and Peter Morrell provided useful comments on an earlier version of this manuscript. This work was supported by National Science Foundation (IOS-0922703 to J.R.-I.); United States Department of Agriculture, National Institute of Food and Agriculture (2009-01864 to J.R.-I.), and Academy of Finland (to T.P.).

Literature Cited

Barthakur N. 1974. Temperature differences between two pigmented types of corn plants. Int J Biometeorol. 18:70–75.
Beaumont MA, Balding DJ. 2004. Identifying adaptive genetic divergence among populations from genome scans. Mol Ecol. 13:969–980.
Beaumont MA, Nichols RA. 1996. Evaluating loci for use in the genetic analysis of population structure. Proc R Soc Lond B Biol Sci. 263: 1619–1626.
Bomblies K, et al. 2010. Local-scale patterns of genetic variability, outcrossing, and spatial structure in natural stands of *Arabidopsis thaliana*. PLoS Genet. 6:e1000890.
Bradburd G, Ralph P, Coop G. 2013. Disentangling the effects of geographic and ecological isolation on genetic differentiation. arXiv:1302.3274.
Bradbury PJ, et al. 2007. TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics 23:2633–2635.
Buckler ES, Goodman MM, Holtsford TP, Doebley JF, Sanchez G. 2006. Phylogeography of the wild subspecies of *Zea mays*. Maydica 51: 123–134.
Chalker-Scott L. 1999. Environmental significance of anthocyanins in plant stress responses. Photochem Photobiol. 70:1–9.
Charlesworth D, Charlesworth B, Morgan MT. 1995. The pattern of neutral molecular variation under the background selection model. Genetics 141:1619–1632.
Cheng C, et al. 2012. Ecological genomics of *Anopheles gambiae* along a latitudinal cline in Cameroon: a population resequencing approach. Genetics 190:1417–1432.
Chia J-M, et al. 2012. Capturing extant variation from a genome in flux: maize HapMap II. Nat Genet. 44:803–807.
Clausen J, Keck DD, Hiesey WM. 1940. Experimental studies on the nature of species. I. Effect of varied environments on western North American plants. Washington (DC): Gibson Brothers, Inc.
Cook JP, et al. 2012. Genetic architecture of maize kernel composition in the nested association mapping and inbred association panels. Plant Physiol. 158:824–834.
Coop G, Witonsky D, Di Rienzo A, Pritchard JK. 2010. Using environmental correlations to identify loci underlying local adaptation. Genetics 185: 1411–1423.
Cutter AD, Payseur BA. 2013. Genomic signatures of selection at linked sites: unifying the disparity among species. Nat Rev Genet. 14: 262–274.
De A, Durrett R. 2007. Stepping-stone spatial structure causes slow decay of linkage disequilibrium and shifts the site frequency spectrum. Genetics 176:969–981.
Dobzhansky TG. 1970. Genetics of the evolutionary process. New York: Columbia University Press.

Doebley J, Goodman MM, Stuber CW. 1987. Pattens of isozyme variation between maize and Mexican annual teosinte. Econ Bot. 41:234–246.
Doyle GG. 1994. Inversions and list of inversions available. In: Freeling M, Walbot V, editors. The maize handbook. New York: Springer-Verlag. p. 346–349.
Earl DA, vonHoldt BM. 2012. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conserv Genet Resour. 4:359–361.
Eckert AJ, et al. 2010a. Patterns of population structure and environmental associations to aridity across the range of loblolly pine (*Pinus taeda* L., Pinaceae). Genetics 185:969–982.
Eckert AJ, et al. 2010b. Back to nature: ecological genomics of loblolly pine (*Pinus taeda*, Pinaceae). Mol Ecol. 19:3789–3805.
Endler JA. 1980. Natural selection on color patterns in *Poecilia reticulata*. Evolution 34:76–91.
Etges WJ, Levitan M. 2004. Palaeoclimatic variation, adaptation and biogeography of inversion polymorphisms in natural populations of *Drosophila robusta*. Biol J Linn Soc. 81:395–411.
Excoffier L, Hofer T, Foll M. 2009. Detecting loci under selection in a hierarchically structured population. Heredity 103:285–298.
Excoffier L, Lischer HEL. 2010. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Mol Ecol Resour. 10:564–567.
Falush D, Stephens M, Pritchard JK. 2003. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. Genetics 164:1567–1587.
Fang Z, et al. 2012. Megabase-scale inversion polymorphism in the wild ancestor of maize. Genetics 191:883–894.
Feder JL, Nosil P. 2010. The efficacy of divergence hitchhiking in generating genomic islands during ecological speciation. Evolution 64: 1729–1747.
Flint-Garcia SA, et al. 2005. Maize association population: a high resolution platform for quantitative trait locus dissection. Plant J. 44:1054–1064.
Foll M, Gaggiotti O. 2008. A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. Genetics 180:977.
Ford HD, Ford EB. 1930. Fluctuation in numbers, and its influence on variation, in *Melitaea aurinia*, Rott. (Lepidoptera). Ecol Entomol. 78: 345–352.
Fournier-Level A, et al. 2011. A map of local adaptation in *Arabidopsis thaliana*. Science 334:86–89.
Fukunaga K, et al. 2005. Genetic diversity and population structure of teosinte. Genetics 169:2241–2254.
Fumagalli M, et al. 2011. Signatures of environmental genetic adaptation pinpoint pathogens as the main selective pressure through human evolution. PLoS Genet. 7:e1002355.
Galinat WC. 1967. Plant habit and the adaptation of corn. Massachusetts Agri Exp Station Bulletin. 565:1–16.
Ganal MW, et al. 2011. A large maize (*Zea mays* L.) SNP genotyping array: development and germplasm genotyping, and genetic mapping to compare with the B73 reference genome. PLoS One 6:e28334.
Gerke JP, Edwards JW, Guill KE, Ross-Ibarra J, McMullen MD. 2013. The genomic impacts of drift and selection for hybrid performance in maize. arXiv 1307.7313.
Gore MA, et al. 2009. A first-generation haplotype map of maize. Science 326:1115.
Goudet J. 2005. HIERFSTAT, a package for R to compute and test hierarchical F-statistics. Mol Ecol Notes. 5:184–186.
Grant PR, Grant BR. 2002. Unpredictable evolution in a 30-year study of Darwin's finches. Science 296:707–711.
Green RE, et al. 2010. A draft sequence of the Neandertal genome. Science 328:710–722.
Gregory TR, et al. 2007. Eukaryotic genome size databases. Nucleic Acids Res. 35:D332–D338.
Gusev A, et al. 2009. Whole population, genome-wide mapping of hidden relatedness. Genome Res. 19:318–326.
Hancock AM, et al. 2011a. Adaptations to climate-mediated selective pressures in humans. PLoS Genet. 7:e1001375.
Hancock AM, et al. 2011b. Adaptation to climate across the *Arabidopsis thaliana* genome. Science 334:83–86.
Hijmans RJ, Guarino L, Cruz M, Rojas E. 2001. Computer tools for spatial analysis of plant genetic resources data: 1. DIVA-GIS. Plant Genet Resour Newsl. 127:15–19.
Hufford MB, Bilinski P, Pyhäjärvi T, Ross-Ibarra J. 2012. Teosinte as a model system for population and ecological genomics. Trends Genet. 28: 606–615.
Hufford MB, Martínez-Meyer E, Gaut BS, Eguiarte LE, Tenaillon MI. 2012. Inferences from the historical distribution of wild and domesticated maize provide ecological and evolutionary insight. PLoS One 7: e47659.
Hufford MB, et al. 2012. Comparative population genomics of maize domestication and improvement. Nat Genet. 44:808–811.
Hufford MB, et al. 2013. The genomic signature of crop-wild introgression in maize. PLoS Genet. 9:e1003477.
Hung H-Y, et al. 2012. *ZmCCT* and the genetic basis of daylength adaptation underlying the post-domestication spread of maize. Proc Natl Acad Sci U S A. 109:11068–11069.
Huynh LY, Maney DL, Thomas JW. 2010. Chromosome-wide linkage disequilibrium caused by an inversion polymorphism in the white--throated sparrow (*Zonotrichia albicollis*). Heredity 106:537–546.
Iltis HH, Doebley JF. 1980. Taxonomy of Zea (Gramineae). II. Subspecific categories in the *Zea mays* complex and a generic synopsis. Am J Bot. 67:994–1004.
Jackson D, Hake S. 1999. Control of phyllotaxy in maize by the *abphyl1* gene. Development 126:315–323.
Kato Y, Miwa K, Takano J, Wada M, Fujiwara T. 2009. Highly boron deficiency-tolerant plants generated by enhanced expression of NIP5;1, a boric acid channel. Plant Cell Physiol. 50:58–66.
Kato YTA. 1975. Cytological studies of maize (*Zea mays* L.) and teosinte (*Zea mexicana* (Schrader) Kuntze) in relation to their origin and evolution. Massachusetts Agri Exp Station Bulletin. 635:1–185.
Keller SR, Levsen N, Olson MS, Tiffin P. 2012. Local adaptation in the flowering time gene network of balsam poplar, *Populus balsamifera* L. Mol Biol Evol. 29:3143–3152.
Kirkpatrick M. 2010. How and why chromosome inversions evolve. PLoS Biol. 8:e1000501.
Kirkpatrick M, Barton N. 2006. Chromosome inversions, local adaptation and speciation. Genetics 173:419–434.
Lauter N, Gustus C, Westerbergh A, Doebley J. 2004. The inheritance and evolution of leaf pigmentation and pubescence in teosinte. Genetics 167:1949–1959.
Le Corre V, Kremer A. 2012. The genetic differentiation at quantitative trait loci under local adaptation. Mol Ecol. 21:1548–1566.
Lewontin RC, Krakauer J. 1973. Distribution of gene frequency as a test of the theory of the selective neutrality of polymorphisms. Genetics 74: 175–195.
Li YF, Costello JC, Holloway AK, Hahn MW. 2008. "Reverse ecology" and the power of population genomics. Evolution 62:2984–2994.
Loveless MD, Hamrick JL. 1984. Ecological determinants of genetic structure in plant populations. Annu Rev Ecol Syst. 15:65–95.
Lowry DB, Rockwood RC, Willis JH. 2008. Ecological reproductive isolation of coast and inland races of *Mimulus guttatus*. Evolution 62: 2196–2214.
Lowry DB, Willis JH. 2010. A widespread chromosomal inversion polymorphism contributes to a major life-history transition, local adaptation, and reproductive isolation. PLoS Biol. 8:e1000500.
Mano Y, Omori F, Takeda K. 2012. Construction of intraspecific linkage maps, detection of a chromosome inversion, and mapping of QTL for

constitutive root aerenchyma formation in the teosinte *Zea nicaraguensis*. Mol Breed. 29:137–146.
Mathieson I, McVean G. 2012. Differential confounding of rare and common variants in spatially structured populations. Nat Genet. 44: 243–246.
Medina RF, Reyna SM, Bernal JS. 2012. Population genetic structure of a specialist leafhopper on Zea: likely anthropogenic and ecological determinants of gene flow. Entomol Exp Appl. 142:223–235.
Meyerowitz EM, Pruitt RE. 1985. *Arabidopsis thaliana* and plant molecular genetics. Science 229:1214–1218.
Moeller DA, Tenaillon MI, Tiffin P. 2007. Population structure and its effects on patterns of nucleotide polymorphism in teosinte (*Zea mays* ssp. *parviglumis*). Genetics 176:1799–1809.
Moeller DA, Tiffin P. 2008. Geographic variation in adaptation at the molecular level: a case study of plant immunity genes. Evolution 62: 3069–3081.
Oksanen J, et al. 2011. vegan: Community Ecology Package, R package version 1.17-6.
Olsson K, Ågren J. 2002. Latitudinal population differentiation in phenology, life history and flower morphology in the perennial herb *Lythrum salicaria*. J Evol Biol. 15:983–996.
Patterson N, Price AL, Reich D. 2006. Population structure and eigenanalysis. PLoS Genet. 2:e190.
Platt A, et al. 2010. The scale of population structure in *Arabidopsis thaliana*. PLoS Genet. 6:e1000843.
Poggio L, Rosato M, Chiavarino AM, Naranjo CA. 1998. Genome size and environmental correlations in maize (*Zea mays* ssp. *mays*, Poaceae). Ann Bot. 82:107–115.
R Core Team. 2012. R: a language and environment for statistical computing. Vienna (Austria): R Foundation for Statistical Computing.
Rosenberg NA. 2004. DISTRUCT: a program for the graphical display of population structure. Mol Ecol Notes. 4:137–138.
Ross-Ibarra J, Tenaillon M, Gaut BS. 2009. Historical divergence and gene flow in the genus *Zea*. Genetics 181:1399–1413.
Saghai-Maroof MA, Soliman KM, Jorgensen RA, Allard RW. 1984. Ribosomal DNA spacer-length polymorphisms in barley: Mendelian inheritance, chromosomal location, and population dynamics. Proc Natl Acad Sci U S A. 81:8014–8018.
Salvi S, et al. 2007. Conserved noncoding genomic sequences associated with a flowering-time quantitative trait locus in maize. Proc Natl Acad Sci U S A. 104:11376–11381.
Sanchez PA, Palm CA, Buol SW. 2003. Fertility capability soil classification: a tool to help assess soil quality in the tropics. Geoderma 114:157–185.
Scheet P, Stephens M. 2006. A fast and flexible statistical model for large-scale population genotype data: applications to inferring missing genotypes and haplotypic phase. Am J Hum Genet. 78:629–644.
Schnable PS, et al. 2009. The B73 maize genome: complexity, diversity, and dynamics. Science 326:1112–1115.
Sheldon E, Ferl R, Fedoroff N, Curtis Hannah L. 1983. Isolation and analysis of a genomic clone encoding sucrose synthetase in maize: evidence for two introns in *Sh*. Mol Gen Genet. 190:421–426.
Siol M, Wright SI, Barratt SCH. 2010. The population genomics of plant adaptation. New Phytol. 188:313–332.
Smith JSC, Goodman MM, Lester RN. 1981. Variation within teosinte. I. Numerical analysis of morphological data. Econ Bot. 35:187–203.
Smith JSC, Goodman MM, Stuber CW. 1984. Variation within teosinte. III. Numerical analysis of allozyme data. Econ Bot. 38:97–113.
St. Onge KR, Palmé AE, Wright SI, Lascoux M. 2012. Impact of sampling schemes on demographic inference: an empirical study in two species with different mating systems and demographic histories. G3 2:803–814.
Städler T, Haubold B, Merino C, Stephan W, Pfaffelhuber P. 2009. The impact of sampling schemes on the site frequency spectrum in nonequilibrium subdivided populations. Genetics 182:205–216.
Storz JF. 2005. Using genome scans of DNA polymorphism to infer adaptive population divergence. Mol Ecol. 14:671–688.
Studer A, Zhao Q, Ross-Ibarra J, Doebley J. 2011. Identification of a functional transposon insertion in the maize domestication gene *tb1*. Nat Genet. 43:1160–1163.
Ting YC. 1964. Chromosomes of maize-teosinte hybrids. Cambridge (MA): The Bussey Institution of Harvard University.
Toomajian C, et al. 2006. A nonparametric test reveals selection for rapid flowering in the *Arabidopsis* genome. PLoS Biol. 4:e137.
Turner TL, Bourne EC, Von Wettberg EJ, Hu TT, Nuzhdin SV. 2010. Population resequencing reveals local adaptation of *Arabidopsis lyrata* to serpentine soils. Nat Genet. 42:260–263.
Turner TL, Von Wettberg EJ, Nuzhdin SV. 2008. Genomic analysis of differentiation between soil types reveals candidate genes for local adaptation in *Arabidopsis lyrata*. PLoS One 3:e3183.
van Heerwaarden J, et al. 2011. Genetic signals of origin, spread, and introgression in a large sample of maize landraces. Proc Natl Acad Sci U S A. 108:1088–1092.
van Heerwaarden J, et al. 2010. Fine scale genetic structure in the wild ancestor of maize (*Zea mays* ssp. *parviglumis*). Mol Ecol. 19: 1162–1173.
Wang H, Bennetzen JL. 2012. Centromere retention and loss during the descent of maize from a tetraploid ancestor. Proc Natl Acad Sci U S A. 109:21004–21009.
Weber A, et al. 2007. Major regulatory genes in maize contribute to standing variation in teosinte (*Zea mays* ssp. *parviglumis*). Genetics 177:2349–2359.
Weber AL, et al. 2008. The genetic architecture of complex traits in teosinte (*Zea mays* ssp. *parviglumis*): new evidence from association mapping. Genetics 180:1221–1232.
Weir BS, Cockerham CC. 1984. Estimating F-statistics for the analysis of population structure. Evolution 38:1358–1370.
Wilkes HG. 1967. Teosinte: the closest relative of maize. Cambridge (MA): The Bussey Institution of Harvard University.
Wolfgruber TK, et al. 2009. Maize centromere structure and evolution: sequence analysis of centromeres 2 and 5 reveals dynamic loci shaped primarily by retrotransposons. PLoS Genet. 5: e1000743.
Wright S, Dobzhansky T. 1946. Genetics of natural populations. XII. Experimental reproduction of some of the changes caused by natural selection in certain populations of *Drosophila pseudoobscura*. Genetics 31:125–156.
Yang RC. 1998. Estimating hierarchical F-statistics. Evolution 950–956.

Associate editor: Bill Martin