UC Riverside

UC Riverside Previously Published Works

Title

The Genomic Signature of Crop-Wild Introgression in Maize

Permalink

https://escholarship.org/uc/item/4c01g7zb

Authors

Hufford, Matthew B
Lubinsky, Pesach
Pyhä̈ jä rvi, Tanya
et al.

Publication Date

2013-05-01

Peer reviewed

The Genomic Signature of Crop-Wild Introgression in Maize

Matthew B. Hufford[1], Pesach Lubinksy[2], Tanja Pyhäjärvi[1], Michael T. Devengenzo[1], Norman C. Ellstrand[3], Jeffrey Ross-Ibarra[1,4]*

1 Department of Plant Sciences, University of California Davis, Davis, California, United States of America, 2 Foreign Agricultural Service, United States Department of Agriculture, Washington, D.C., United States of America, 3 Department of Botany and Plant Sciences, University of California Riverside, Riverside, California, United States of America, 4 Genome Center and Center for Population Biology, University of California Davis, Davis, California, United States of America

Abstract

The evolutionary significance of hybridization and subsequent introgression has long been appreciated, but evaluation of the genome-wide effects of these phenomena has only recently become possible. Crop-wild study systems represent ideal opportunities to examine evolution through hybridization. For example, maize and the conspecific wild teosinte *Zea mays* ssp. *mexicana* (hereafter, *mexicana*) are known to hybridize in the fields of highland Mexico. Despite widespread evidence of gene flow, maize and *mexicana* maintain distinct morphologies and have done so in sympatry for thousands of years. Neither the genomic extent nor the evolutionary importance of introgression between these taxa is understood. In this study we assessed patterns of genome-wide introgression based on 39,029 single nucleotide polymorphisms genotyped in 189 individuals from nine sympatric maize-*mexicana* populations and reference allopatric populations. While portions of the maize and *mexicana* genomes appeared resistant to introgression (notably near known cross-incompatibility and domestication loci), we detected widespread evidence for introgression in both directions of gene flow. Through further characterization of these genomic regions and preliminary growth chamber experiments, we found evidence suggestive of the incorporation of adaptive *mexicana* alleles into maize during its expansion to the highlands of central Mexico. In contrast, very little evidence was found for adaptive introgression from maize to *mexicana*. The methods we have applied here can be replicated widely, and such analyses have the potential to greatly inform our understanding of evolution through introgressive hybridization. Crop species, due to their exceptional genomic resources and frequent histories of spread into sympatry with relatives, should be particularly influential in these studies.

Introduction

Hybridization and subsequent introgression have long been appreciated as agents of evolution. Adaptations can be transferred through these processes upon secondary contact of uniquely adapted populations or species, in many instances producing the variation necessary for further diversification [1]. Early considerations of adaptive introgression discussed its importance in the context of both domesticated and wild species [2,3], viewing both anthropogenic disturbance and naturally heterogeneous environments as ideal settings for hybridization. More recently, studies of adaptation through introgression have focused primarily on wild species ([4,5] but see [6,7]). Well-studied examples include increased hybrid fitness of Darwin's finches following environmental changes that favor beak morphology intermediate to that found in extant species [8,9] and the introgression of traits related to herbivore resistance [10] and drought escape [11] in species of wild sunflower [12,13]. Molecular and population genetic analyses have also clearly identified instances of adaptive introgression across species at individual loci, including examples such as the RAY locus controlling floral morphology and outcrossing rate in

groundsels [14]) and the optix gene controlling wing color in mimetic butterflies [15,16]. Despite long-standing interest in introgression, however, genome-wide analyses are rare and have been primarily conducted in model systems [17–22].

Studies of natural introgression in cultivated species have been limited in genomic scope and have largely ignored the issue of historical adaptive introgression, focusing instead on contemporary transgene escape and/or the evolution of weediness [23–27]. One notable exception is recent work documenting introgression between different groups of cultivated rice in genomic regions containing loci involved in domestication [19,28–30]. Few studies, however, have investigated the potential for introgression to transfer adaptations between crops and natural populations of their wild relatives post-domestication. Subsequent to domestication, most crops spread from centers of origin into new habitats, often encountering locally adapted populations of their wild progenitors and closely related species (*e.g.*, [31–33]). These crop expansions provide compelling opportunities to study evolution through introgressive hybridization.

Here, we use a SNP genotyping array to investigate the genomic signature of gene flow between cultivated maize and its wild

Author Summary

Hybridization and introgression have been shown to play a critical role in the evolution of species. These processes can generate the diversity necessary for novel adaptations and continued diversification of taxa. Previous research has suggested that not all regions of a genome are equally permeable to introgression. We have conducted one of the first genome-wide assessments of patterns of reciprocal introgression in plant populations. We found evidence that suggests domesticated maize received adaptation to highland conditions from a wild relative, teosinte, during its spread to the high elevations of central Mexico. Gene flow appeared asymmetric, favoring teosinte introgression into maize, and was widespread across populations at putatively adaptive loci. In contrast, genomic regions near known domestication and cross-incompatibility loci appeared particularly resistant to introgression in both directions of gene flow. Crop-wild study systems should play an important role in future studies of introgression due to their well-developed genomic resources and histories of reciprocal gene flow during crop expansion.

relative *Zea mays* ssp. *mexicana* (hereafter, *mexicana*) and examine evidence for adaptive introgression. Maize was domesticated approximately 9,000 BP in southwest Mexico from the lowland teosinte taxon *Zea mays* ssp. *parviglumis* (hereafter, *parviglumis*; [34–36]). Following domestication, maize spread to the highlands of central Mexico [34,37], a migration that involved adaptation to thousands of meters of changing elevation and brought maize to substantially cooler ($\sim7^{\circ}$C change in annual temperature) and drier (~300 mm change in annual precipitation) climes [38]. During this migration maize came into sympatry with *mexicana*, a highland teosinte that diverged from *parviglumis* $\sim60,000$ BP [39].

Convincing morphological evidence for introgression between maize and *mexicana* has been reported [40,41], and traits putatively involved in adaptation to the cooler highland environment such as dark-red and highly-pilose leaf sheaths [42,43] are shared between *mexicana* and highland maize landraces [40,44]. These shared morphological features could suggest adaptive introgression [45] but could also reflect parallel or convergent adaptation to highland climate or retention of ancestral traits [46]. Though hybrids are frequently observed, phenological isolation due to flowering time differences [40,47] and cross-incompatibility loci [48–50] are thought to limit the extent of introgression, particularly acting as barriers to maize pollination of *mexicana*. Experimental estimates of maize-*mexicana* pollination success (*i.e.*, production of hybrid seed) are quite low, ranging from <1–2% depending on the direction of the cross [51,52]. Nevertheless, theory suggests that alleles received through hybridization can persist and spread despite such barriers to gene exchange, particularly when they prove adaptive [53,54].

Molecular analyses over the last few decades have provided increasingly strong evidence for reciprocal introgression between *mexicana* and highland maize landraces. Early work identified multiple allozyme alleles common in highland Mexican maize and *mexicana* but rare in closely related taxa or maize outside of the highlands [55]. Likewise, sequencing of the putative domestication locus *barren stalk1* (*ba1*) revealed a haplotype unique to *mexicana* and highland Mexican maize [56]. Multiple studies have found further support for bidirectional gene flow and have estimated that ~2–10% of the genome of highland maize is derived from *mexicana* [34,57] and 4–8% of the *mexicana* genome is derived from maize [58]. A more recent study including several hundred markers

revealed that admixture with *mexicana* may approach 20% in highland Mexican maize [36].

Similar to introgression studies in many other plant species (*e.g.*, [31,59–62]), morphological and molecular studies have provided only rough estimates of the extent of introgression between *mexicana* and maize. Little is known regarding genome-wide patterns in the extent and directionality of gene flow. A genomic picture of introgression could greatly expand our understanding of evolution through hybridization, revealing how particular alleles, genes and genomic regions are disproportionately shaped by and/or resistant to these processes [63,64]. Additionally, assessment of introgression in crop species during post-domestication expansion can provide insight into the genetic architecture of adaptation to newly encountered abiotic and biotic conditions. Here, we provide the most in-depth analysis to date of the genomic extent and directionality of introgression in sympatric collections of maize and its wild relative, *mexicana*, based on genome-wide single nucleotide polymorphism (SNP) data. We find evidence for pervasive yet asymmetric gene flow in sympatric populations. Across the genome, several regions introgressed from *mexicana* into maize are shared across most populations, while little consistency in introgression is observed in gene flow in the opposite direction. These data, combined with analysis of environmental associations and a growth chamber experiment, suggest that maize colonization of highland environments in Mexico may have been facilitated by adaptive introgression from local *mexicana* populations.

Results

Polymorphism and Differentiation

To assess the extent of hybridization and introgression we collected nine sympatric population pairs of maize and *mexicana* and one allopatric *mexicana* population from across the highlands of Mexico (Table S1; Figure 1) and genotyped 189 individuals for 39,029 SNPs (see Materials and Methods). Genotype data at the same loci were obtained from Chia *et al.* [65] for a reference allopatric maize population. Average expected heterozygosity (H_E), percent polymorphic loci (%P), and the proportion of privately segregating sites were higher in maize than *mexicana* (t-test, p≤0.012 for all comparisons, Table S2), likely influenced by the absence of *mexicana* from the discovery panel used to develop the genotyping platform [66]. However, substantial variation in diversity was observed across populations within taxa (*e.g.*, %P ranged from 52–88% in maize and from 44–79% in *mexicana* (Table S2)) and meaningful comparisons can be made at this level. Our analysis of diversity identified the Ixtlan maize population as an extreme outlier, containing 31% fewer polymorphic markers than any other maize population. Discussion with farmers during our collection revealed that Ixtlan maize was initially a commercial variety whose seed had been replanted for a number of generations. Excluding this population, diversity in *mexicana* populations varied much more substantially than in maize (*e.g.*, variance in %P across *mexicana* populations was 7-fold higher; Table S2)

At the population level, summary statistics of diversity and differentiation were consistent with sympatric gene flow (*i.e.*, local gene flow based on current plant distributions) between maize and *mexicana* (Figure 2). First, %P was positively correlated between sympatric population pairs ($R^2 = 0.65$; p = 0.016; Figure 2A), though this trend could reflect local conditions affecting diversity in both taxa rather than gene flow. Second, in a subset of populations, the proportion of shared polymorphisms was higher (Figure 2B) and pairwise differentiation (F_{ST}) was lower (Figure 2C)
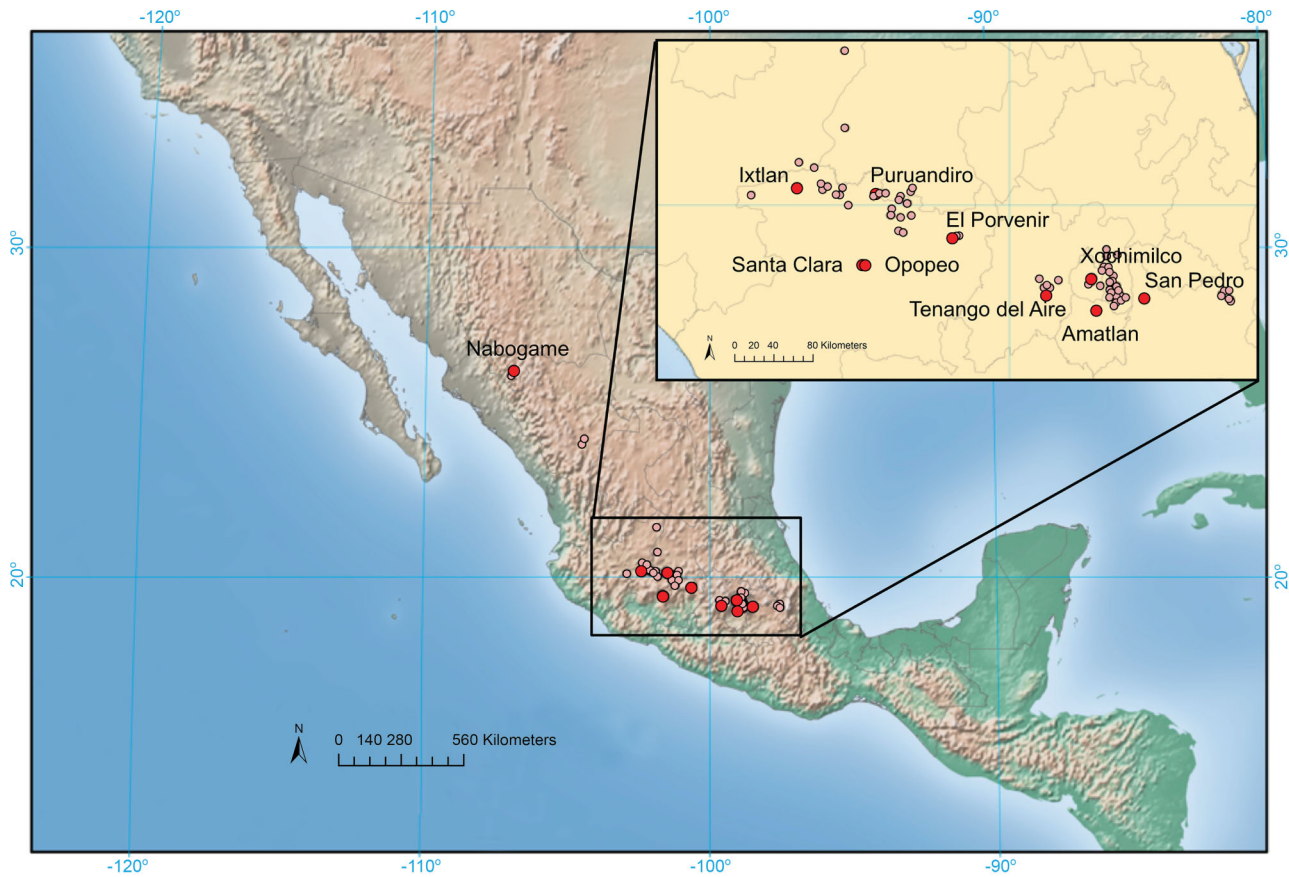
Figure 1. Map of collection sites. Light red dots indicate known *mexicana* populations and larger, dark red dots indicate populations included in the current study.
doi:10.1371/journal.pgen.1003477.g001

between sympatric population pairs than in allopatric comparisons. Finally, an individual-based STRUCTURE analysis assuming two groups (K = 2) revealed strong membership of reference allopatric individuals of maize and *mexicana* in their appropriate groups (96% and 99% respectively), yet appreciable admixture in sympatric populations (Figure 2D). Four recent hybrids were identified (3 *mexicana* and 1 maize) with <60% membership in their respective groups. STRUCTURE analysis also indicated that gene flow was asymmetric, with more highland maize germplasm derived from *mexicana* (19% versus 12% of *mexicana* germplasm from maize). Assignment at higher K values continued to indicate admixture in *mexicana* populations but not in maize, suggesting that gene flow from *mexicana* into maize may have been more ancient (Figure S1). Consistent with this interpretation, median values of the *f*3 statistic [67] for SNPs genome-wide were negative or zero for 8 of 9 sympatric maize populations (Figure S2); only the Ixtlan maize population showed a positive median *f*3 signifying a lack of admixture. Collectively, these population-level summaries are suggestive of historical gene flow from *mexicana* into maize and, in a subset of populations, of ongoing sympatric gene flow from maize into *mexicana*.

Variation in Introgression Levels across the Genome

Meaningful information regarding the evolutionary significance of introgression can often be obscured in population-level summaries. However, the large number of SNPs in our data set allowed us to assess variation in the extent of introgression across the genome. We made use of two complementary methods. First,

we employed the hidden Markov model of HAPMIX [68] to infer ancestry of chromosomal segments along the genomes of individuals from maize and *mexicana* populations through comparison to reference allopatric populations. Subsampling of the reference allopatric populations (see Materials and Methods) revealed considerable signal of introgression in the maize reference panel, particularly in low recombination regions of the genome near centromeres (correction for this signal is illustrated in Figure 3 and Figure S3). While this signal could represent genuine introgression predating allopatry, it could also indicate potential false positives in genomic regions with high linkage disequilibrium or less data. We therefore added a complementary analysis using the linkage model of STRUCTURE [69,70] to conduct site-by-site assignment across the genomes of *mexicana* and maize. Because STRUCTURE takes allele frequencies across all populations into account during assignment, the approach is robust to potential deviations of individual reference populations from ancestral frequencies.

Both methods allowed quantification of introgression along the genome for individual samples. Rather than investigate every putative introgression, however, we focused further analyses on genomic regions with a high frequency of introgression, requiring an average of one chromosome or 50% assignment to the opposite taxon per individual in a given population (Figure 3; Figure S3; referred to as ''introgressed regions'' hereafter). Approximately 19.1% and 9.8% of the genome met this criterion in the HAPMIX and STRUCTURE scans respectively for *mexicana* introgression into maize. In the opposite direction, we observed lower
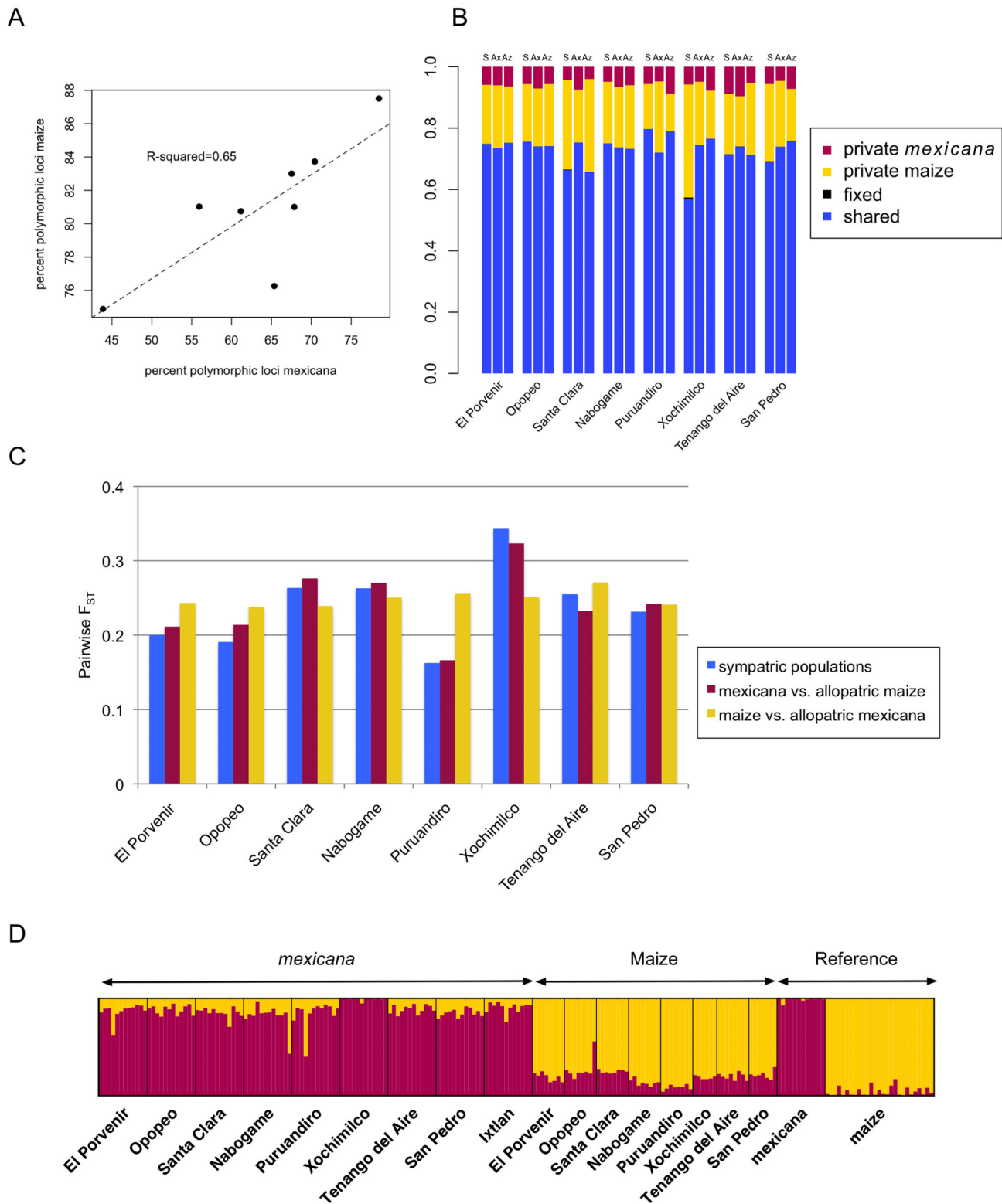
Figure 2. Population-level polymorphism and differentiation. (A) Correlation of percent polymorphic loci in sympatric populations of *mexicana* and maize. (B) Proportion of shared and privately segregating polymorphisms in *mexicana* and maize and fixed differences between taxa. Letters above bars indicate sympatric maize/*mexicana* comparisons (S), maize from a given population versus allopatric *mexicana* (Ax) and *mexicana* from a given population versus allopatric maize (Az). (C) Pairwise differentiation (F_{ST}) in sympatric and allopatric comparisons of *mexicana* and maize. (D) Bar plot of assignment proportions from STRUCTURE analysis at K = 2 for *mexicana* (maroon) and maize (gold) individuals. The Ixtlan maize population was excluded from this figure and the STRUCTURE analysis.
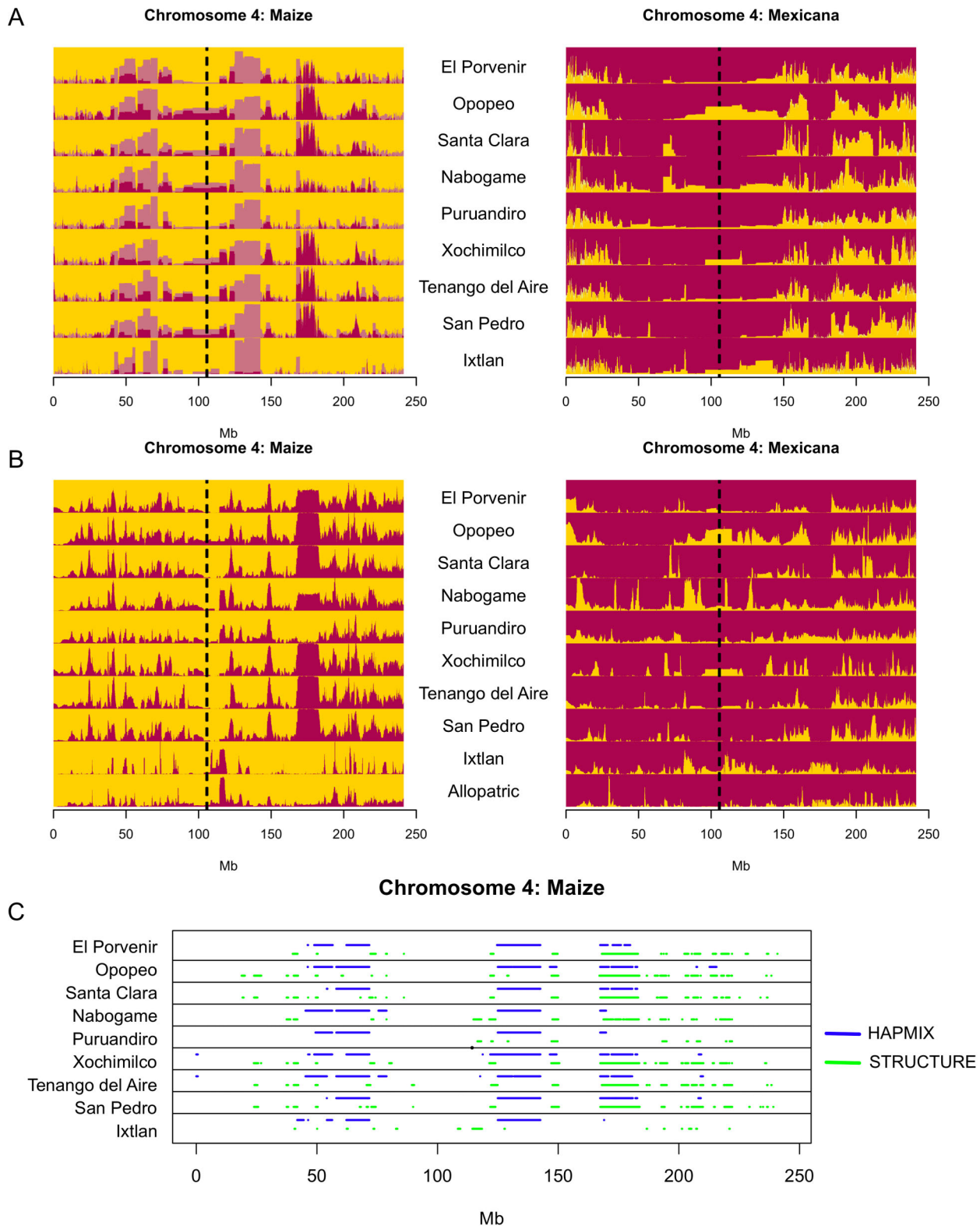doi:10.1371/journal.pgen.1003477.g002

Figure 3. Detection of introgression across chromosome 4. (A) Stacked bar plots of the HAPMIX introgression scan across sympatric populations. Population labels are indicated between plots for maize (gold) and *mexicana* (maroon). Lighter colors indicate introgression initially detected in each population and darker colors show these values after subtracting introgression proportions from jackknife samples of the allopatric reference populations that may be due to false positives. A dotted black line indicates the position of the centromere. (B) Stacked bar plots of the STRUCTURE introgression scan across sympatric populations. The *mexicana* group is indicated by maroon and the maize group is indicated by gold. The y-axis for each population in (A) and (B) indicates the average admixture proportion across individuals. (C) Genomic regions in maize populations showing greater than 50% membership in *mexicana*.
doi:10.1371/journal.pgen.1003477.g003

proportions at this threshold (11.4% in the case of HAPMIX and 9.2% using STRUCTURE), corroborating asymmetric gene flow favoring *mexicana* introgression into maize. Both scans showed a disproportionate number of introgressed regions shared across populations in *mexicana*-to-maize gene flow. Roughly 50% of regions introgressed from *mexicana* into maize were shared across seven or more populations in the HAPMIX scan, whereas only 4% of introgressed regions had this level of sharing from maize into *mexicana*; similar asymmetry was observed using STRUCTURE (12% versus <1%).

By comparing composite likelihood scores from HAPMIX across individuals within each population, we were able to characterize relative times since admixture (see Materials and Methods). We observed qualitative differences between maize and *mexicana*. The likelihood of the admixture time parameter began to decrease markedly after an average of 83 generations in *mexicana* populations, whereas the decrease in maize was much more gradual and did not occur until after an average of 174 generations (Figure S4; averages exclude Ixtlan) suggesting older introgression from *mexicana* into maize. A notable exception to this trend was observed in the Ixtlan sympatric population pair, where the maize population was likely derived in the recent past from a commercial variety and introgression appeared to be more recent from *mexicana* into maize (Figure S4).

For further population genetic characterization, we focused on the subset of introgressed regions identified in both the HAPMIX and STRUCTURE scans, an approach that should be robust to the individual assumptions of the two methods. These regions spanned an average of 3.6% of the genome in the case of *mexicana*-to-maize introgression and 3.2% for maize-to-*mexicana* introgression (Figure 3C; Figure S3). As expected, differentiation between sympatric maize and *mexicana* was reduced in these introgressed regions in both directions of gene flow (mean 25% reduction of F_{ST} *mexicana*-to-maize, 33% reduction maize-to-*mexicana*, t-test, p<0.001 for all population-level comparisons of introgressed vs. non-introgressed regions in both directions of gene flow). Introgressed regions also showed more shared and fewer fixed and private SNPs (Table S3), as well as longer tracts of identity by state (IBS) between maize and *mexicana* (t-test, p<<0.001). Consistent with these results, diversity in introgressed regions was generally different from non-introgressed regions in the recipient taxon and instead comparable to diversity in non-introgressed regions in the taxon of origin (Table S3).

In total, we identified nine regions of introgression from *mexicana* to maize found by both methods and present in ≥7 sympatric population pairs (Table S4). Three of these shared regions of introgression span the centromeres of chromosomes 5, 6, and 10 (Figure S3), suggesting that maize from the highlands of Mexico may in fact harbor *mexicana* centromeric or pericentromeric sequence. No such shared introgressions were found in the opposite direction of gene flow (maize into *mexicana*).

Finally, we characterized regions of the genome notably lacking evidence of introgression. We refer to regions with ≤5% probability of introgression confirmed by both scans in ≥7 populations as being resistant to introgression (Figure S5). In both directions of gene flow, we found these genomic regions to have elevated differentiation, decreased diversity, fewer shared variants, more fixed differences, and a higher number of privately segregating SNPs in the opposite taxon (Table S3).

Evaluating Evidence for Adaptive Introgression

Two non-mutually exclusive hypotheses of adaptive introgression can be readily discerned for gene flow between *mexicana* and maize: 1) as its natural habitat was transformed, *mexicana* received

maize alleles conferring adaptation to the agronomic setting and 2) as it diffused to the highlands of central Mexico from the lowlands of southwest Mexico, maize received alleles conferring highland adaptation from *mexicana*, which was already adapted to these conditions. To evaluate evidence for the first hypothesis we gauged enrichment of 484 candidate domestication genes [71] in regions of introgression. We hypothesized that if maize donated alleles adaptive for the agronomic setting to *mexicana*, we would detect enrichment of domestication loci in regions introgressed from maize into *mexicana*. However, compared to the rest of the genome, introgressed regions in both directions of gene flow harbored significantly fewer domestication candidates (permutation test, p≤0.001), while regions resistant to introgression showed an excess of domestication candidates (permutation test, p = 0.121 maize to *mexicana*, p = 0.008 *mexicana* to maize; Figure S5). For example, two well-characterized domestication genes affecting branching architecture, *grassy tillers1* (*gt1*; [72,73]) and *teosinte branched1* (*tb1*; [74]) showed very little evidence of introgression (Figure S5). Introgression also appeared to be rare from maize into *mexicana* across much of the short arm of chromosome 4, a span that includes the domestication loci *teosinte glume architecture1* (*tga1*; [75]), *sugary1* (*su1*; [76]) and *brittle endosperm2* (*bt2*; [76]) and the well characterized pollen-pistil incompatibility locus *teosinte crossing barrier1* (*tcb1*; [48]) that serves as a hybridization barrier between maize and *mexicana* (Figure S5). These results suggest selection against introgression at loci that contribute to domestication and reproductive isolation.

Several lines of evidence support the hypothesis that maize received introgression conferring highland adaptation from *mexicana*. Across the nine shared introgressed regions, five contained long stretches (>300 kb) of zero diversity across seven populations, implying a common introgressed haplotype (Figure S6). Given that these regions only have 5–15 SNPs, however, higher-density genotyping might resolve additional haplotypes. Additionally, we used the method of Coop et al. [77] to detect associations of population allele frequencies with 76 environmental variables (see Materials and Methods). Environmental variables were reduced in dimensionality to four principal components that captured 95% of environmental variation. We found that loci associated with the second principal component (loaded primarily by temperature seasonality) were significantly enriched (permutation test, p = 0.017) in genomic regions introgressed from *mexicana* into maize, but no significant enrichment was observed in regions introgressed from maize into *mexicana*. We then compared the nine regions of introgression found in ≥7 populations of maize to QTL for anthocyanin content and leaf macrohairs (putatively adaptive traits under highland conditions) identified in a previous study from a cross between *parviglumis* (lowland teosinte) and *mexicana* (highland teosinte) [42]. Six of the introgressed regions overlapped with five of the six genomic regions with QTL detected for these traits.

Two of the shared introgressions that overlapped with QTL are of particular interest due to their previous characterization. One of these, on chromosome 4, overlaps with QTL for both pigment intensity and macrohairs [42], and maps to the same position as a recently identified putative inversion polymorphism showing significant differentiation between *parviglumis* and *mexicana* ([78]; Figure 4A). The second region, on chromosome 9, overlaps with a QTL for macrohairs [42] and includes the *macrohairless1* (*mhl1*) locus [79] that promotes macrohair formation on the leaf blade and sheath of maize (Figure 4B). The two lowest elevation maize populations in our study (Puruandiro and Ixtlan) showed a conspicuous lack of introgression in these two genomic regions (Figure 4A and 4B). Analysis of pairwise differentiation (F_{ST})

between these populations and two populations showing fixed introgression in the two genomic regions (Opopeo and San Pedro; Figure 4A and 4B) revealed substantial differentiation: the region on chromosome 4 contained the only fixed SNP differences genome-wide (Puruandiro/Ixtlan versus Opopeo/San Pedro) and a SNP in the region on chromosome 9 was an extreme F_{ST} outlier. To explore the potential phenotypic effects of these genomic regions we conducted growth chamber experiments including ten maize plants from each of these four populations. Under temperature and day-length conditions typical of the highlands of Mexico (see Materials and Methods), the leaf sheaths of plants from populations where introgression was detected in the two genomic regions had 21-fold more macrohairs (t-test, p = 0.0002; Figure 4C and 4D), and showed greater pigmentation (t-test, p = $6E^{-06}$; Figure 4C and 4D). Introgressed plants were also ~25 cm taller (t-test, p = $6E^{-06}$; Figure 4D), a finding consistent with adaptation to highland conditions and potentially associated with increased fitness. No significant difference in plant height was observed in a separate experiment under lowland conditions (t = test, p = 0.51), and a significant interaction was observed between introgression status and environmental treatment (ANOVA, F = 4.151, p = 0.045), with a disproportionate increase in plant height under lowland conditions in populations lacking introgression (Figure S7).

Contribution of mexicana to Modern Maize Lines

While our scans for introgression clearly indicated that *mexicana* has made genomic contributions to maize landraces in the highlands of Mexico, the broader contribution of *mexicana* to modern maize lines remained unclear. Our HAPMIX and STRUCTURE analyses had low power to detect introgression distributed broadly in maize (see Discussion). Therefore, to assess potential ancestral contribution of *mexicana* to modern maize, we evaluated patterns of IBS between *mexicana*, *parviglumis* [78] and a global diversity panel of 279 modern maize lines [80,81] using the program GERMLINE ([82]; Figure 5, Figure S8 and S9). Substantial IBS was found between *mexicana* and modern lines at a number of genomic locations. To assess whether this IBS merely reflected shared ancestral haplotypes, we compared IBS between modern maize and *parviglumis* to IBS between modern maize and *mexicana* on a site-by-site basis, identifying regions in which various maize groups distinguished by Flint-Garcia et al. [81] showed stronger IBS with *mexicana* relative to *parviglumis* (see Materials and Methods; Figure 5A; Figure S8). As each of the groups identified by Flint-Garcia have distinct evolutionary histories, it is possible that *mexicana* contributed differentially to the founders of each group. For example, the tropical-subtropical, non-stiff-stalk, and mixed groups showed more genomic regions with stronger IBS with *mexicana* (versus *parviglumis*) than found in the stiff-stalk, popcorn, and sweetcorn groups (~31% of sites with greater IBS with *mexicana* in the first group versus ~23% in the latter group; Figure 5B and 5C).

Discussion

Despite known pre-zygotic and phenological barriers to hybridization between maize and *mexicana* [47–50], we have found evidence consistent with substantial reciprocal introgression. Based on our population genetic analyses, several observations regarding the nature of this gene flow can be made: 1) Gene flow appears to be ongoing and asymmetric, favoring *mexicana* introgression into maize. 2) Gene flow from *mexicana* into maize is generally older than gene flow in the opposite direction. 3) Haplotype diversity in nine genomic regions of *mexicana*-into-maize introgression shared

across ≥7 populations suggests single, ancient introgressions followed by spread across the Mexican highlands. 4) Introgression from *mexicana* into maize is restricted at domestication loci but enriched at loci putatively involved in highland adaptation. 5) Genomic regions of *mexicana*/maize IBS within a global diversity panel of maize hint at a possible broader contribution of *mexicana* to modern improved maize.

Several of these observations are in line with previous research. For example, the asymmetric gene flow we detect from *mexicana* to maize is consistent with findings of substantially higher pollination success in this direction [51]. Asymmetric gene flow would also be expected based on phenology: in Mexico, maize typically flowers earlier than *mexicana* [47] and pollen shed in both taxa precedes silking (female flowering). Therefore, when maize silks are receptive, *mexicana* could potentially be shedding pollen, whereas when *mexicana* silks are receptive, maize tassels are more likely to be senescent. Under these conditions, F1 progeny would be more likely to have a maize seed parent and a teosinte pollen parent and subsequent inadvertent planting of F1's in maize fields would bias the direction of gene flow.

Our data also provide support for previous assertions that shared morphological features between *mexicana* and maize represent adaptations derived from *mexicana* [45] rather than from maize [41]. For example, we have found significant environmental correlations in genomic regions of *mexicana*-to-maize introgression. We have also observed that overlap with QTL and fine-mapped loci for highland *Zea* traits (*e.g.*, leaf sheath macrohairs and pigmentation) are predominantly found in the direction of *mexicana* to maize gene flow. Two such regions, on chromosomes 4 and 9, showed particularly strong evidence of introgression. Moreover, these genomic regions of introgression were more common in higher elevation maize populations in our sample, and maize populations with and without introgression in these regions showed differential morphology and greater plant height (a proxy for fitness) when grown under highland conditions. In contrast, we found little evidence of adaptive introgression in the opposite direction of gene flow. For example, domestication loci appeared resistant to gene flow from maize into *mexicana*, contradicting previous suggestions that gene flow from maize may have been required for *mexicana* to adapt to an agronomic setting [41]. Instead it appears likely that *mexicana*, like other wild teosintes [83], was a ruderal species adapted to open and disturbed environments even before the transformation of its natural habitat by maize cultivation.

Our detection of haplotype sharing between *mexicana* and a diverse panel of modern maize is consistent with previous findings suggesting the spread of introgressed *mexicana* haplotypes in maize outside of the highlands of Mexico [71]. Both the STRUCTURE and HAPMIX methods we used to identify regions of introgression would likely not detect introgression found ubiquitously in modern maize. Widespread *mexicana* introgression into maize would result in poor resolution between reference populations of these taxa in the HAPMIX analysis, and extensive haplotype sharing across maize and *mexicana* would result in a weak signature of introgression in STRUCTURE. Further analysis of representative panels of *mexicana*, *parviglumis* and maize haplotypes at greater marker density should help clearly distinguish *mexicana* from *parviglumis* haplotypes and determine whether *mexicana* haplotypes are indeed widespread in maize.

While our results are consistent with previous research and the historical spread of maize, our power to detect introgression may be limited for a number of reasons. First, our analysis conservatively focused on regions of introgression identified by two independent methods and shared across individuals within
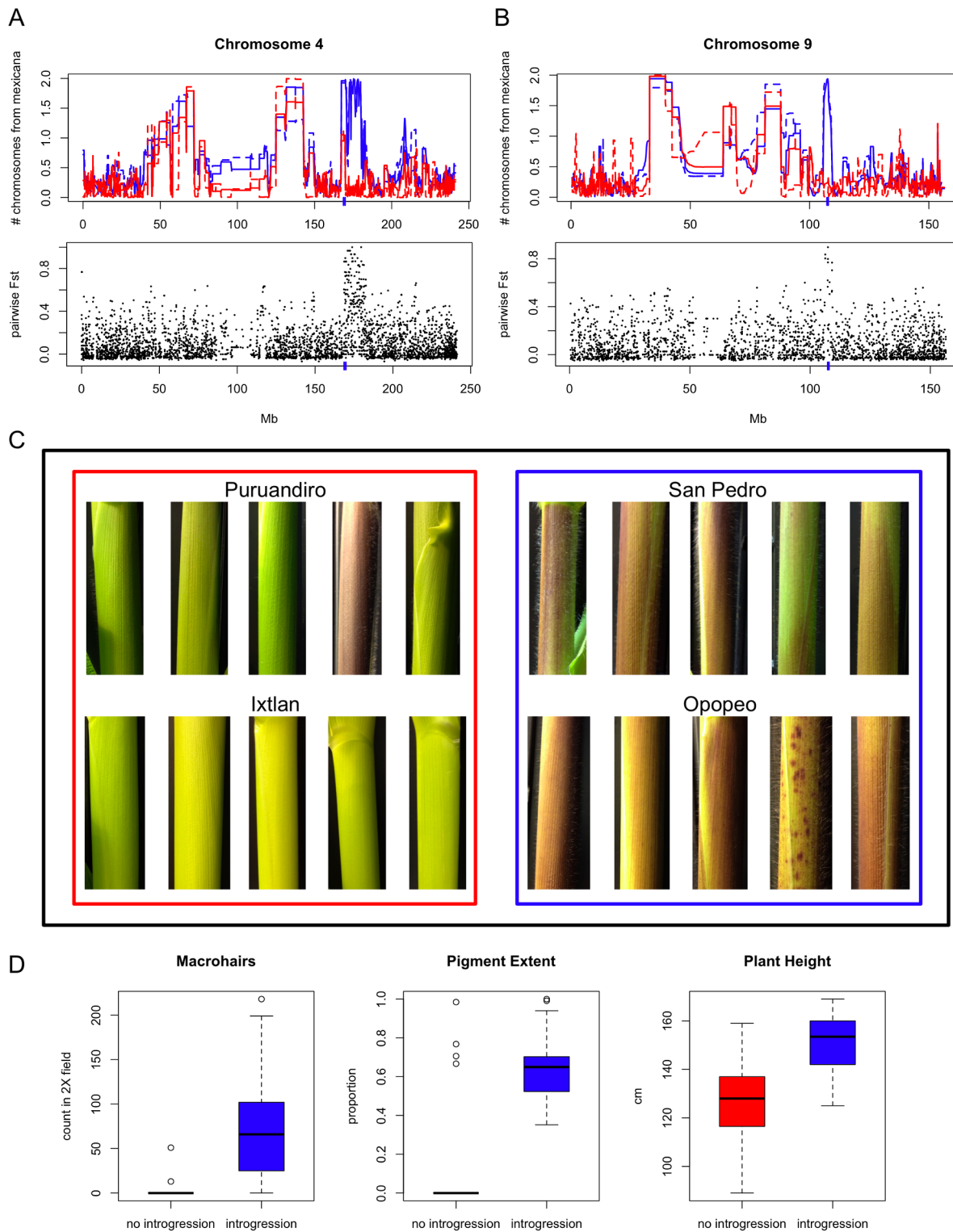
Figure 4. Growth chamber experiment. (A) Region of *mexicana*-to-maize introgression on chromosome 4 (indicated by blue hash on x-axis) shared across seven populations. Above: patterns of introgression seen in San Pedro (blue solid line) and Opopeo (blue dashed line) versus Puruandiro (red solid line) and Ixtlan (red dashed line). Below: pairwise F_{ST} of San Pedro/Opopeo (with introgression) versus Puruandiro/Ixtlan (without introgression). (B) Region of *mexicana*-to-maize introgression on chromosome 9 (indicated by blue hash on x-axis) shared across seven populations. Populations and comparisons are as in Figure 4A. (C) Five leaf sheaths from each of four maize populations grown under highland conditions. (D) Distribution of maize trait values (macrohairs, pigment extent and plant height at 50 days) from growth chamber experiment emulating highland conditions in populations with and without introgression.
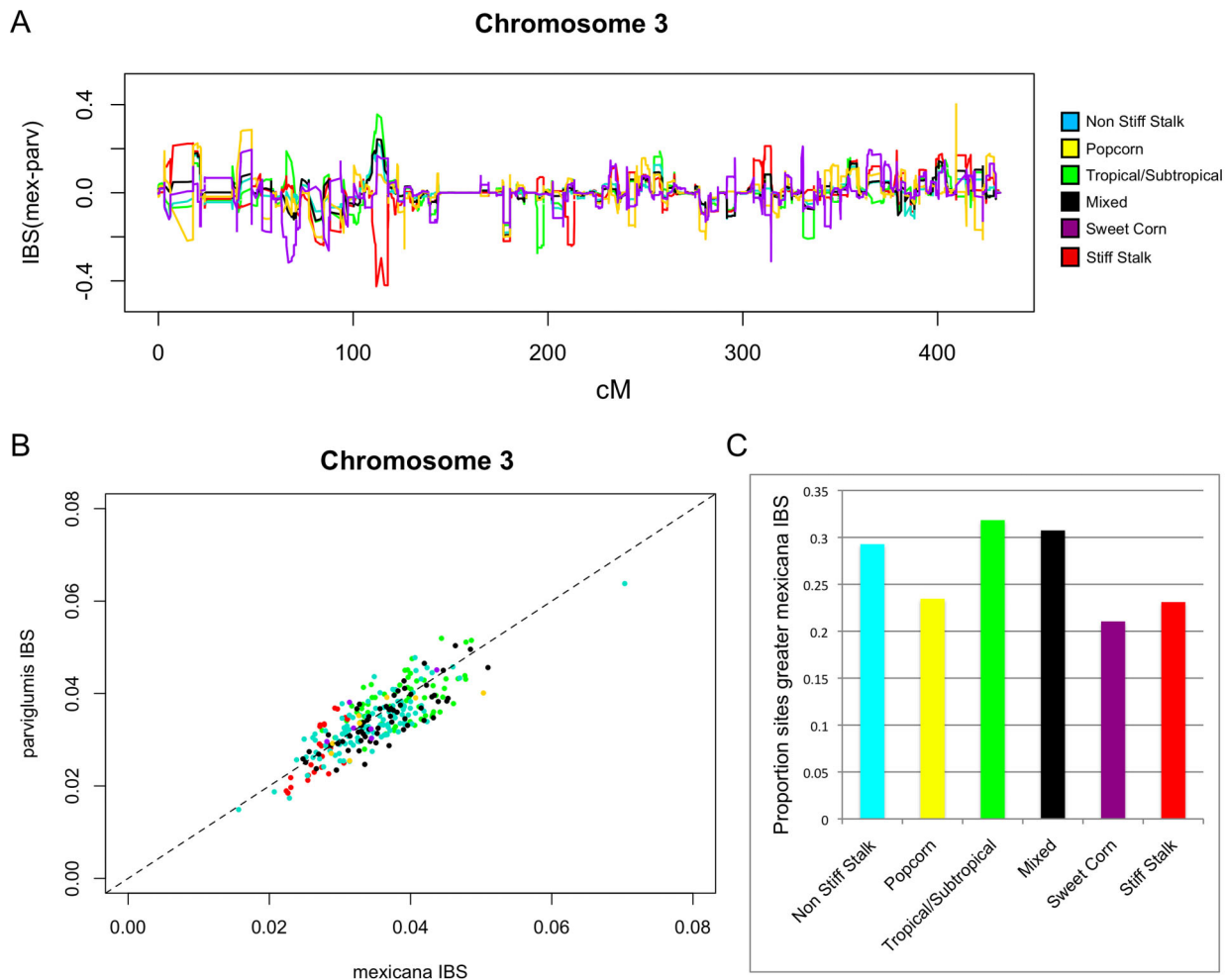doi:10.1371/journal.pgen.1003477.g004

Figure 5. Contribution of *mexicana* germplasm to a global maize diversity panel. (A) The difference between the average IBS proportion with *mexicana* individuals and the average IBS proportion with *parviglumis* individuals calculated across six groups of modern maize lines identified by Flint-Garcia et al. [81] in the maize association population. Positive values indicate greater IBS with *mexicana*. (B) Average IBS across chromosome 3 in each line in the maize association population compared to both *mexicana* and *parviglumis*. The one-to-one line is indicated by the dashed line. Colors are as in Figure 5A. (C) The proportion of sites across the genome showing greater IBS with *mexicana* than with *parviglumis* for each of the six maize association population groups.
doi:10.1371/journal.pgen.1003477.g005

populations, undoubtedly missing a number of genuine instances of more limited gene flow. Second, our markers were ascertained in a panel consisting entirely of maize. In addition to inflating the diversity of maize relative to *mexicana*, this ascertainment scheme likely limited our ability to distinguish among *mexicana* haplotypes and thus to detect local introgression from *mexicana* into maize. Third, the resolution of our data was on average one SNP per 80 kb, which could result in a bias toward detection of more recent introgression and introgression in low recombination regions of the genome. Finally, *mexicana* only rarely occurs allopatric from maize [40], and most populations have likely experienced gene flow at some point in time, thus complicating estimation of ancestral *mexicana* haplotypes and allele frequencies.

Many aspects of *mexicana*'s contribution to highland adaptation in maize remain to be resolved. While our growth chamber experiment was suggestive of adaptive introgression, the loci conferring these traits are still ambiguous. Repetition of these experiments with *mexicana*/lowland maize near-isogenic introgression lines will be necessary to bolster the case for adaptive introgression. Additionally, a particularly interesting comparison

can be made between highland maize in central Mexico, a geographic region sympatric with *mexicana*, and highland maize in the Andes of South America where no inter-fertile wild *Zea* species can be found. Future research should address whether highland adaptation in South American maize occurred in parallel to maize from Mexico [37] or whether pre-adapted highland maize was transported through Central America as some have suggested [84].

The potential for adaptive introgression during crop expansion is of course not limited to maize. Data from several crops (*e.g.*, rice [19,85], barley [86,87], common bean [88], and wheat [32,89]) suggest defined centers of origin within a broader distribution of wild relatives. The distributions of these crop-wild pairs span continents and a wide range of environments, and many are known to hybridize (for a review, see [24]). The methods we have applied here to maize and *mexicana* can therefore be replicated widely, perhaps revealing unexpected aspects of crop evolution and providing insight regarding the genetic architecture of local adaptation based on conserved regions of introgression.

Crops and related wild taxa can also be seen more broadly as models for the study of evolution through hybridization. If crops are viewed as human-facilitated invasive species, clear connections can be made to theoretical work on introgression during invasion and range expansion. For example, our finding of asymmetric gene flow from *mexicana* into maize is consistent with simulations showing that invaders should receive much higher levels of introgression from local species than occurs in the opposite direction due to differences in population density at the time of invasion [90,91]. Theoretical research has also explored the divergence threshold for successful hybridization and introgression [53,92]. Crop expansions are ideal systems to test such predictions because, as ancient agriculturalists moved crops away from their centers of origin, these domesticates came into sympatry with relatives spanning a range of divergence times. For example, *parviglumis*, the progenitor of maize, has a divergence time from *mexicana* estimated at 60,000 years, from other members of the genus on the order of 100,000–300,000 years, and from the outgroup *Tripsacum dactyloides* of approximately 1 million years [39]. While *parviglumis* is currently physically isolated from these taxa and likely was at the time of domestication [38], maize has subsequently come into sympatry with virtually all of its close relatives, providing extensive opportunities for hybridization. These newly-formed hybrid zones can be seen as testing grounds of the fitness of hybrids across a range of divergence and opportunities to study the evolution of barriers to hybridization.

Materials and Methods

Sample Collection and Genotyping

Samples were collected from nine sympatric population pairs of *mexicana* and maize that spanned the known distribution of *mexicana* in Mexico, as well as a single allopatric population of *mexicana* (Table S1; Figure 1). Seed samples from 12 maternal individuals per *mexicana* population (N = 120) were selected for genotyping. A single kernel was also sampled from each of 6–8 maize ears collected from sympatric maize fields (N = 69). The tenth kernel down from the tip of each ear was chosen to help control for potential variation in outcrossing rate along the ear. Seeds were treated with fungicide, germinated on filter paper and grown in standard potting mix to the five-leaf stage. Freshly harvested leaf tips were stored at −80°C overnight and lyophilized for 48 hours. Tissue was then homogenized with a Mini-Beadbeater-8 (BioSpec Products, Inc., Bartlesville, OK, USA) and DNA was isolated using a modified CTAB protocol [93]. Purity of DNA isolations was determined with a NanoDrop spectrophotometer (NanoDrop Technologies, Inc., Wilmington, DE, USA). Samples with 260:280 ratios ≥1.8 were deemed acceptable for genotyping. Concentrations of DNA isolations were determined with a Wallac VICTOR2 fluorescence plate reader (Perkin-Elmer Life and Analytical Sciences, Torrance, CA, USA) using the Quant-iT Picogreen dsDNA Assay Kit (Invitrogen, Grand Island, NY, USA). Single nucleotide polymorphism genotypes were generated using the Illumina MaizeSNP50 Genotyping BeadChip platform and were clustered separately for the two taxa based on the default algorithm of the GenomeStudio Genotyping Module v1.0 (Illumina Inc., San Diego, CA, USA). Clustering for each SNP in each taxon was subsequently inspected and manually adjusted. Of the total of 56,110 markers contained on the chip, 39,029 SNPs that were polymorphic within the entire sample of maize and *mexicana* and contained less than 10% missing data in both taxa were used for further analysis.

Diversity Analyses

Observed (H_O) and expected (H_E) heterozygosities were summarized for each taxon in each sympatric population pair using the "genetics" package in R [94]. Polymorphisms were further characterized as shared, fixed, or segregating privately within one of each pair of sympatric populations using the sharedPoly program of the libsequence C++ library [95]. Pairwise differentiation between populations (F_{ST}) was calculated based on the method of Weir and Cockerham [96] using custom R scripts and the "hierfstat" package of R [97]. The $f3$ statistic for identification of admixture [67] was calculated using a custom R script.

Detecting Introgression

To characterize patterns of introgression across the genome in each population we used two complementary methods: 1) Identification of ancestry across chromosomal segments with the hidden Markov model approach of HAPMIX [68]; and 2) A site-by-site analysis of assignment probabilities using the Bayesian linkage model in the program STRUCTURE [69,70]. For both HAPMIX and STRUCTURE analyses, we used a subset of 38,262 SNPs anchored in a genetic map based on the Intermated B73×Mo17 (IBM) population of maize ([66]; J.P. Gerke et al., unpublished data). The IBM population has been widely used for genetic map development and for determining the genetic architecture of complex traits in maize [98].

Patterns of introgression were assessed using the program HAPMIX by comparing unphased data from putatively admixed individuals from our sympatric populations to phased data from reference ancestral populations. To represent ancestral *mexicana* haplotypes, we chose a population near the town of Amatlán, Morelos state, Mexico that is currently allopatric to maize. An Americas-wide sample of maize landraces collected largely outside the distribution of *mexicana* was chosen as the maize reference population [65]. In order to assess putative introgression and/or false positives in these reference populations, we removed each individual and evaluated introgression through comparison to remaining reference samples using a jackknife approach. Evidence for introgression was assessed in both putatively admixed and reference individuals using HAPMIX as described below.

Initial estimates of ancestry proportions for HAPMIX models were based on a previous admixture analysis of *mexicana* and highland Mexican maize (~20% introgression of *mexicana* into maize and ~10% introgression of maize into *mexicana*; [36]). The number of generations since the time of admixture was varied from 1–5000 and the maximum likelihood across individuals in a population was used to compare relative time since admixture on a population-by-population basis (Figure S4). Subsequent analyses of HAPMIX output were based on introgression estimates from the highest likelihood run.

Prior to analysis in STRUCTURE, SNP data were phased using the program fastPHASE (version 1.4.0; [99]). Because STRUCTURE does not account for linkage disequilibrium (LD) due to physical linkage, SNPs were grouped into haplotypes separated by at least 5 kb. After grouping, our data set consisted of 20,035 loci with an average of 3.92 alleles per locus across all sympatric and reference allopatric individuals. We ran the linkage model in STRUCTURE with 5,000 steps of admixture burn-in, a total burn-in of 10,000 steps, and 100,000 subsequent steps retained for analysis. Convergence along the chain and consistency across replicate runs were assessed to ensure an adequate number of steps were included in the analysis. Assignment was carried out for K = 2 groups (*i.e.*, maize and *mexicana*) for each chromosome

separately. Probability of assignment was summarized locus by locus across individuals from each population for each taxon.

Local Adaptation at Introgressed Loci

To identify SNPs associated with environmental variables, we employed the association method of BAYENV [77], using a covariance matrix of allele frequencies estimated with 10,000 random SNPs to control for population structure. Seventy-six climatic and soil variables were summarized as four principal components that captured 95% of the variance among *mexicana* populations. BAYENV was run five times with 1,000,000 iterations for each SNP. A given SNP was considered a candidate if its Bayes factor was consistently in the 95th percentile across all five independent runs and its average Bayes factor was in the 99th percentile. Enrichment of significant SNPs in introgressed regions was determined based on bootstrap resampling for each environmental PC.

Haplotype Sharing

Analyses of haplotype sharing/identity by state between *mexicana*, *parviglumis*, and modern maize lines were conducted using the program GERMLINE [82] with haplotypes generated by the program fastPHASE [99] from samples of *parviglumis* [78] and modern maize [80]. Shared haplotypes were identified with a seed of identical genotypes at five SNPs that were extended until mismatch. Analyses were then based on segments with a minimum size of 3 cM.

Growth Chamber Experiment

Ten seeds were germinated from each of four maize populations showing little evidence of introgression (Ixtlan and Puruandiro) or fixed introgression (Opopeo and San Pedro) at two loci (one on chromosome 4 and one on chromosome 9; Table S4) putatively linked to highland adaptation [42,78] and showing little evidence of false positives in our reference populations. Plants were grown under highland conditions with 12.5 hours of light at an intensity of 680 μmol/m^2*s, a daytime temperature of 23°C and a nighttime temperature of 11°C. Daytime relative humidity was set at 60% and nighttime relative humidity at 80%. Height measurements were taken at 15, 30, and 50 days. Pigment extent was measured on the second leaf sheath from the top of the plant as the proportion of the total sheath showing pigment. Macrohairs were also measured on this leaf sheath as the total count one third of the way down from the leaf blade within the field of a dissecting microscope at 2× magnification. In order to contrast plant height from our highland treatment to those under conditions more comparable to the lowlands of western Mexico, we conducted a separate growth chamber experiment with a daytime temperature of 32°C and a nighttime temperature of 25°C and measured plant height at 30 days. All other conditions were identical to those of the highland treatment.

Supporting Information

Figure S1 Bar plot of assignment proportions from STRUCTURE analysis at K = 2-K = 10 for *mexicana* and maize individuals. The Ixtlan maize population was excluded from this figure and the STRUCTURE analysis.
(PDF)

Figure S2 Distribution of *f3* statistic in genome-wide SNPs genotyped in nine maize populations that are sympatric with *mexicana*.
(PDF)

Figure S3 HAPMIX and STRUCTURE plots of introgression for each chromosome. Colors and axes are as in Figure 3.
(PDF)

Figure S4 Likelihood plots across generations since admixture for each population for both *mexicana* (A) and maize (B).
(PDF)

Figure S5 Proportion of populations showing resistance to introgression across each chromosome for maize-to-*mexicana* (A) and *mexicana*-to-maize (B) introgression. Thirteen well-known domestication loci (red) and three characterized pollen-pistil cross-incompatibility loci (blue) are indicated with dashed lines and labeled above the plots.
(PDF)

Figure S6 Evidence supporting reduced haplotype diversity in shared regions of *mexicana*-to-maize introgression. (A) Expected heterozygosity in introgressed (blue) and non-introgressed (red) individuals in shared regions. Expected heterozygosity plotted across chromosomes 4 (B) and 9 (C) in introgressed individuals. Regions of introgression shared across populations are indicated by a red dashed line.
(PDF)

Figure S7 Plant height at 30 days in maize populations with (blue) and without (red) introgression at loci depicted in Figure 4 under highland and lowland conditions. Confidence interval is +/−1 standard error.
(PDF)

Figure S8 Identity by State (IBS) of modern maize lines with *mexicana* and *parviglumis* across each chromosome. All plots are as in Figure 5B.
(PDF)

Figure S9 The difference between IBS modern maize/*mexicana* and IBS modern maize/*parviglumis* across each chromosome. All plots are as in Figure 5A. Dashed lines indicate genomic regions of *mexicana* introgression into highland Mexican maize conserved across ≥7 populations.
(PDF)

Table S1 Sampling information for *mexicana* and maize populations.
(PDF)

Table S2 Summaries of diversity across *mexicana* and maize populations. H_E = expected heterozygosity, $\%P$ = percent polymorphic loci, H_O = observed heterozygosity, F_{IS} = inbreeding coefficient calculated as $(H_E - H_O)/H_E$.
(PDF)

Table S3 Population genetic summaries from introgressed regions and regions resisting introgression. Parameters are as in Table S2. Significant differences (permutation or t-test, p<0.05) between introgressed and non-introgressed regions are indicated as bold values.
(XLSX)

Table S4 Genomic coordinates of shared introgression regions.
(PDF)

Author Contributions

Conceived and designed the experiments: MBH PL NCE JR-I. Performed the experiments: MBH PL MTD. Analyzed the data: MBH TP. Contributed reagents/materials/analysis tools: NCE JR-I. Wrote the paper: MBH JR-I.

References

1. Stebbins GL (1959) The role of hybridization in evolution. P Am Philos Soc 103: 231–251.
2. Anderson E (1948) Hybridization of the habitat. Evolution 2: 1–9.
3. Anderson E, Stebbins GL (1954) Hybridization as an evolutionary stimulus. Evolution 8: 378–388.
4. Arnold ML (2004) Transfer and origin of adaptations through natural hybridization: Were Anderson and Stebbins right? Plant Cell 16: 562–570.
5. Arnold ML, Martin NH (2010) Hybrid fitness across time and habitats. Trends Ecol Evol 25: 530–536.
6. Arnold ML (2004) Natural hybridization and the evolution of domesticated, pest and disease organisms. Mol Ecol 13: 997–1007.
7. Ellstrand NC, Heredia SM, Leak-Garcia JA, Heraty JM, Burger JC, et al. (2010) Crops gone wild: evolution of weeds and invasives from domesticated ancestors. Evol Appl 3: 494–504.
8. Grant BR, Grant PR (1996) High survival of Darwin's finch hybrids: Effects of beak morphology and diets. Ecology 77: 500–509.
9. Grant PR, Grant BR (2010) Conspecific versus heterospecific gene exchange between populations of Darwin's finches. Philos T Roy Soc B 365: 1065–1076.
10. Whitney KD, Randell RA, Rieseberg LH (2006) Adaptive introgression of herbivore resistance traits in the weedy sunflower *Helianthus annuus*. Am Nat 167: 794–807.
11. Whitney KD, Randell RA, Rieseberg LH (2010) Adaptive introgression of abiotic tolerance traits in the sunflower *Helianthus annuus*. New Phytol 187: 230–239.
12. Heiser CB (1951) Hybridization in the annual sunflowers - *Helianthus-annuus*X*H-debilis* var-*cucumerifolius*. Evolution 5: 42–51.
13. Kim SC, Rieseberg LH (1999) Genetic architecture of species differences in annual sunflowers: Implications for adaptive trait introgression. Genetics 153: 965–977.
14. Kim M, Cui M-L, Cubas P, Gillies A, Lee K, et al. (2008) Regulatory genes control a key morphological and ecological trait transferred between species. Science 322: 1116–1119.
15. Dasmahapatra KK, Walters JR, Briscoe AD, Davey JW, Whibley A, et al. (2012) Butterfly genome reveals promiscuous exchange of mimicry adaptations among species. Nature 487: 94–98.
16. Pardo-Diaz C, Salazar C, Baxter SW, Merot C, Figueiredo-Ready W, et al. (2012) Adaptive introgression across species boundaries in *Heliconius* butterflies. PLoS Genet 8: e1002752. doi:10.1371/journal.pgen.1002752
17. Durand EY, Patterson N, Reich D, Slatkin M (2011) Testing for ancient admixture between closely related populations. Mol Biol Evol 28: 2239–2252.
18. Green RE, Krause J, Briggs AW, Maricic T, Stenzel U, et al. (2010) A draft sequence of the Neandertal genome. Science 328: 710–722.
19. He Z, Zhai W, Wen H, Tang T, Wang Y, et al. (2011) Two evolutionary histories in the genome of rice: the roles of domestication genes. PLoS Genet 7: e1002100. doi:10.1371/journal.pgen.1002100
20. Kulathinal RJ, Stevison LS, Noor MAF (2009) The genomics of speciation in *Drosophila*: diversity, divergence, and introgression estimated using low-coverage genome sequencing. PLoS Genet 5: e1000550. doi:10.1371/journal.pgen.1000550
21. Reich D, Green RE, Kircher M, Krause J, Patterson N, et al. (2010) Genetic history of an archaic hominin group from Denisova Cave in Siberia. Nature 468: 1053–1060.
22. Sun Y, Corcoran P, Menkis A, Whittle CA, Andersson SGE, et al. (2012) Large-scale introgression shapes the evolution of the mating-type chromosomes of the filamentous ascomycete *Neurospora tetrasperma*. PLoS Genet 8: e1002820. doi:10.1371/journal.pgen.1002820
23. Arnaud JF, Viard F, Delescluse M, Cuguen J (2003) Evidence for gene flow via seed dispersal from crop to wild relatives in *Beta vulgaris* (Chenopodiaceae): consequences for the release of genetically modified crop species with weedy lineages. P Roy Soc Lond B Bio 270: 1565–1571.
24. Ellstrand NC (2003) Dangerous Liaisons? When Cultivated Plants Mate with Their Wild Relatives. Baltimore, Maryland: The Johns Hopkins University Press. 244 p.
25. Gutierrez A, Cantamutto M, Poverene M (2011) Persistence of sunflower crop traits and fitness in *Helianthus petiolaris* populations. Plant Biol 13: 821–830.
26. Morrell PL, Williams-Coplin TD, Lattu AL, Bowers JE, Chandler JM, et al. (2005) Crop-to-weed introgression has impacted allelic composition of johnsongrass populations with and without recent exposure to cultivated sorghum. Mol Ecol 14: 2143–2154.
27. Xia HB, Wang W, Xia H, Zhao W, Lu BR (2011) Conspecific crop-weed introgression influences evolution of weedy rice (*Oryza sativa* f. *spontanea*) across a geographical range. PLoS ONE 6: e16189. doi:10.1371/journal.pone.0016189
28. Fujino K, Wu J, Sekiguchi H, Ito T, Izawa T, et al. (2010) Multiple introgression events surrounding the *Hd1* flowering-time gene in cultivated rice, *Oryza sativa* L. Mol Genet Genomics 284: 137–146.
29. McNally KL, Childs KL, Bohnert R, Davidson RM, Zhao K, et al. (2009) Genomewide SNP variation reveals relationships among landraces and modern varieties of rice. P Natl Acad Sci USA 106: 12273–12278.
30. Zhao K, Wright M, Kimball J, Eizenga G, McClung A, et al. (2010) Genomic Diversity and introgression in *O. sativa* reveal the impact of domestication and breeding on the rice genome. PLoS ONE 5: e10780. doi:10.1371/journal.pone.0010780
31. Cornille A, Gladieux P, Smulders MJM, Roldán-Ruiz I, Laurens F, et al. (2012) New insight into the history of domesticated apple: secondary contribution of the European wild apple to the genome of cultivated varieties. PLoS Genet 8: e1002703. doi:10.1371/journal.pgen.1002703
32. Luo MC, Yang ZL, You FM, Kawahara T, Waines JG, et al. (2007) The structure of wild and domesticated emmer wheat populations, gene flow between them, and the site of emmer domestication. Theor Appl Genet 114: 947–959.
33. Myles S, Boyko AR, Owens CL, Brown PJ, Grassi F, et al. (2011) Genetic structure and domestication history of the grape. P Natl Acad Sci USA 108: 3530–3535.
34. Matsuoka Y, Vigouroux Y, Goodman MM, Sanchez GJ, Buckler E, et al. (2002) A single domestication for maize shown by multilocus microsatellite genotyping. P Natl Acad Sci USA 99: 6080–6084.
35. 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. P Natl Acad Sci USA 106: 5019–5024.
36. van Heerwaarden J, Doebley J, Briggs WH, Glaubitz JC, Goodman MM, et al. (2011) Genetic signals of origin, spread, and introgression in a large sample of maize landraces. P Natl Acad Sci USA 108: 1088–1092.
37. Vigouroux Y, Glaubitz JC, Matsuoka Y, Goodman MM, Jesus SG, et al. (2008) Population structure and genetic diversity of New World maize races assessed by DNA microsatellites. Am J Bot 95: 1240–1253.
38. Hufford MB, Martinez-Meyer E, Gaut BS, Eguiarte LE, Tenaillon MI (2012) Past and present distributions of wild and domesticated *Zea mays*: a chance to revisit maize history. PLoS ONE 7: e47659. doi:10.1371/journal.pone.0047659
39. Ross-Ibarra J, Tenaillon M, Gaut BS (2009) Historical divergence and gene flow in the genus *Zea*. Genetics 181: 1397–1409.
40. Wilkes HG (1967) Teosinte: The Closest Relative of Maize. Cambridge, Massachusetts: The Bussey Institute of Harvard University. 159 p.
41. Wilkes HG (1977) Hybridization of maize and teosinte, in Mexico and Guatemala and improvement of maize. Econ Bot 31: 254–293.
42. Lauter N, Gustus C, Westerbergh A, Doebley J (2004) The inheritance and evolution of leaf pigmentation and pubescence in teosinte. Genetics 167: 1949–1959.
43. Barthakur N (1974) Temperature differences between two pigmented types of corn plants. Int J Biometeorol 18: 70–75.
44. Collins GN (1921) Teosinte in Mexico - The closest wild relative of maize is teosinte - a forage plant hitherto known only as an annual. A perennial form discovered in southern Mexico should prove of value to the breeder. J Hered 12: 339–350.
45. Doebley JF (1984) Maize introgression into teosinte - a reappraisal. Ann Mo Bot Gard 71: 1100–1113.
46. Doebley J (1990) Molecular evidence and the evolution of maize. Econ Bot 44: 6–27.
47. Rodriguez JG, Sanchez JJ, Baltazar B, De la Cruz L, Santacruz-Ruvalcaba F, et al. (2006) Characterization of floral morphology and synchrony among *Zea* species in Mexico. Maydica 51: 383–398.
48. Evans MMS, Kermicle JL (2001) *Teosinte crossing barrier1*, a locus governing hybridization of teosinte with maize. Theor Appl Genet 103: 259–265.
49. Kermicle JL, Evans MMS (2010) The *Zea mays* sexual compatibility gene *ga2*: naturally occurring alleles, their distribution, and role in reproductive isolation. J Hered 101: 737–749.
50. Kermicle JL, Taba S, Evans MMS (2006) The *gametophyte-1* locus and reproductive isolation among *Zea mays* subspecies. Maydica 51: 219–225.
51. Baltazar BM, Sanchez-Gonzalez JD, de la Cruz-Larios L, Schoper JB (2005) Pollination between maize and teosinte: an important determinant of gene flow in Mexico. Theor Appl Genet 110: 519–526.
52. Ellstrand NC, Garner LC, Hegde S, Guadagnuolo R, Blancas L (2007) Spontaneous hybridization between maize and teosinte. J Hered 98: 183–187.
53. Barton NH (2001) The role of hybridization in evolution. Mol Ecol 10: 551–568.
54. Mallet J (2005) Hybridization as an invasion of the genome. Trends Ecol Evol 20: 229–237.
55. Doebley J, Goodman MM, Stuber CW (1987) Patterns of isozyme variation between maize and Mexican annual teosinte. Econ Bot 41: 234–246.
56. Gallavotti A, Zhao Q, Kyozuka J, Meeley RB, Ritter M, et al. (2004) The role of *barren stalk1* in the architecture of maize. Nature 432: 630–635.
57. Warburton ML, Wilkes G, Taba S, Charcosset A, Mir C, et al. (2011) Gene flow among different teosinte taxa and into the domesticated maize gene pool. Genet Resour Crop Ev 58: 1243–1261.

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

59. Delplancke M, Alvarez N, Espindola A, Joly H, Benoit L, et al. (2012) Gene flow among wild and domesticated almond species: insights from chloroplast and nuclear markers. Evol Appl 5: 317–329.

60. Huebner S, Guenther T, Flavell A, Fridman E, Graner A, et al. (2012) Islands and streams: clusters and gene flow in wild barley populations from the Levant. Mol Ecol 21: 1115–1129.

61. Li Z, Zou J, Mao K, Lin K, Li H, et al. (2012) Population genetic evidence for complex evolutionary histories of four high altitude juniper species in the Qinghai-Tibetan plateau. Evolution 66: 831–845.

62. Scascitelli M, Whitney KD, Randell RA, King M, Buerkle CA, et al. (2010) Genome scan of hybridizing sunflowers from Texas (*Helianthus annuus* and *H. debilis*) reveals asymmetric patterns of introgression and small islands of genomic differentiation. Mol Ecol 19: 521–541.

63. Baack EJ, Rieseberg LH. (2007) A genomic view of introgression and hybrid speciation. Curr Opin Genet Dev 17: 513–518.

64. Twyford AD, Ennos RA (2012) Next-generation hybridization and introgression. Heredity 108: 179–189.

65. Chia J-M, Song C, Bradbury PJ, Costich D, de Leon N, et al. (2012) Maize HapMap2 identifies extant variation from a genome in flux. Nat Genet 44: 803–807.

66. Ganal MW, Durstewitz G, Polley A, Berard A, Buckler ES, 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. doi:10.1371/journal.pone.0028334

67. Reich D, Kumarasamy T, Patterson N, Price AL, Singh L (2009) Reconstructing Indian population history. Nature 461: 489–494.

68. Price AL, Tandon A, Patterson N, Barnes KC, Rafaels N, et al. (2009) Sensitive detection of chromosomal segments of distinct ancestry in admixed populations. PLoS Genet 5: e1000519. doi:10.1371/journal.pgen.1000519

69. 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.

70. Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155: 945–959.

71. Hufford MB, Xu X, van Heerwaarden J, Pyhäjärvi T, Chia J-M, et al. (2012) Comparative population genomics of maize domestication and improvement. Nat Genet 808–811.

72. Whipple CJ, Kebrom TH, Weber AL, Yang F, Hall D, et al. (2011) *grassy tillers1* promotes apical dominance in maize and responds to shade signals in the grasses. P Natl Acad Sci USA 108: E506–E512.

73. Wills DM, Whipple C, Takuno S, Kursel LE, Shannon LM, et al. (2013) From many, one: Genetic control of prolificacy during maize domestication. arXiv:1303.0882v1. Available at: http://arxiv.org/abs/1303.0882. Accessed 25 March 2013.

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

75. Wang H, Nussbaum-Wagler T, Li BL, Zhao Q, Vigouroux Y, et al. (2005) The origin of the naked grains of maize. Nature 436: 714–719.

76. Whitt SR, Wilson LM, Tenaillon MI, Gaut BS, Buckler ES (2002) Genetic diversity and selection in the maize starch pathway. P Natl Acad Sci USA 99: 12959–12962.

77. Coop G, Witonsky D, Di Rienzo A, Pritchard JK (2010) Using environmental correlations to identify loci underlying local adaptation. Genetics 185: 1411–1423.

78. Pyhäjärvi T, Hufford MB, Mezmouk S, Ross-Ibarra J (2012) Complex patterns of local adaptation in teosinte. arXiv:1208.0634v1. Available at: http://arxiv.org/abs/1208.0634. Accessed 28 February 2013.

79. Moose SP, Lauter N, Carlson SR (2004) The maize *macrohairless1* locus specifically promotes leaf blade macrohair initiation and responds to factors regulating leaf identity. Genetics 166: 1451–1461.

80. Cook JP, McMullen MD, Holland JB, Tian F, Bradbury P, et al. (2012) Genetic architecture of maize kernel composition in the nested association mapping and inbred association panels. Plant Physiol 158: 824–834.

81. Flint-Garcia SA, Thuillet AC, Yu JM, Pressoir G, Romero SM, et al. (2005) Maize association population: a high-resolution platform for quantitative trait locus dissection. Plant J 44: 1054–1064.

82. Gusev A, Lowe JK, Stoffel M, Daly MJ, Altshuler D, et al. (2009) Whole population, genome-wide mapping of hidden relatedness. Genome Res 19: 318–326.

83. Sanchez-Velasquez LR, Ezcurra E, Martinez-Ramos M, Alvarez-Buylla E, Lorente R (2002) Population dynamics of *Zea diploperennis*, an endangered perennial herb: effect of slash and burn practice. J Ecol 90: 684–692.

84. Freitas FO, Bendel G, Allaby RG, Brown TA (2003) DNA from primitive maize landraces and archaeological remains: implications for the domestication of maize and its expansion into South America. J Archaeol Sci 30: 901–908.

85. Molina J, Sikora M, Garud N, Flowers JM, Rubinstein S, et al. (2011) Molecular evidence for a single evolutionary origin of domesticated rice. P Natl Acad Sci USA 108: 8351–8356.

86. Badr A, Muller K, Schafer-Pregl R, El Rabey H, Effgen S, et al. (2000) On the origin and domestication history of barley (*Hordeum vulgare*). Mol Biol Evol 17: 499–510.

87. Morrell PL, Clegg MT (2007) Genetic evidence for a second domestication of barley (*Hordeum vulgare*) east of the Fertile Crescent. P Natl Acad Sci USA 104: 3289–3294.

88. Kwak M, Kami JA, Gepts P (2009) The putative Mesoamerican domestication center of *Phaseolus vulgaris* is located in the Lerma-Santiago Basin of Mexico. Crop Sci 49: 554–563.

89. Heun M, SchaferPregl R, Klawan D, Castagna R, Accerbi M, et al. (1997) Site of einkorn wheat domestication identified by DNA fingerprinting. Science 278: 1312–1314.

90. Currat M, Ruedi M, Petit RJ, Excoffier L (2008) The hidden side of invasions: Massive introgression by local genes. Evolution 62: 1908–1920.

91. Excoffier L, Foll M, Petit RJ (2009) Genetic consequences of range expansions. Annu Rev Ecol Evol S 40: 481–501.

92. Gavrilets S (1997) Hybrid zones with Dobzhansky-type epistatic selection. Evolution 51: 1027–1035.

93. Saghai-Maroof MA, Soliman KM, Jorgensen RA, Allard RW (1984) Ribosomal DNA spacer-length polymorphisms in barley - Mendelian inheritance, chromosomal location, and population-dynamics. P Natl Acad Sci USA 81: 8014–8018.

94. R Development Core Team (2012) R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.

95. Thornton K (2003) libsequence: a C++ class library for evolutionary genetic analysis. Bioinformatics 19: 2325–2327.

96. Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population-structure. Evolution 38: 1358–1370.

97. Goudet J (2005) HIERFSTAT, a package for R to compute and test hierarchical F-statistics. Mol Ecol Notes 5: 184–186.

98. Lee M, Sharopova N, Beavis WD, Grant D, Katt M, et al. (2002) Expanding the genetic map of maize with the intermated B73×Mo17 (IBM) population. Plant Mol Biol 48: 453–461.

99. 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.