Tracking footprints of maize domestication and evidence for a massive selective sweep on chromosome 10

Feng Tian[a], Natalie M. Stevens[a], and Edward S. Buckler IV[a,b,c,1]

[a]Institute for Genomic Diversity and [c]Department of Plant Breeding and Genetics, Cornell University, Ithaca, NY 14850; and [b]United States Department of Agriculture, Agricultural Research Service, Ithaca, NY 14850

Maize domestication is one of the greatest feats of artificial selection and evolution, wherein a weedy plant in Central Mexico was converted through human-mediated selection into the most productive crop in the world. In fact, the changes were so astounding that it took much of the last century to identify modern maize's true ancestor. Through modern genetic studies, the molecular basis of this evolution is being unraveled. Maize's new morphology and adaptation to diverse environments required selection at thousands of loci, and we are beginning to understand the magnitude and rates of these genetic changes. Most of the known major genes have experienced strong selection, but only small regions surrounding the selected genes exhibit substantially reduced genetic diversity. Here, we report the discovery of a large region on chromosome 10 involved in adaptation or domestication that has been the target of strong selection during maize domestication. Unlike previously described regions in the maize genome, 1.1 Mb and >15 genes lost genetic diversity during selection at this region. Finally, the prospects of a detailed understanding of maize evolution are discussed with consideration of both top-down and bottom-up approaches.

> Although man does not cause variability and cannot even prevent it, he can select, preserve, and accumulate the variations given to him by the hand of nature almost in any way that he chooses; and thus he can certainly produce a great result.
>
> Charles Darwin (1)

With its meager ear containing only 2 entwined rows of well-armored kernels, teosinte grows on Mexican hillsides. This grass might easily have been overlooked were it not for its abundant variation, a gift not lost on early agriculturists. Within the last 10,000 years, early Native Americans were able to transform teosinte into a plant whose ears would feed the world. It was a transformation so striking and so complex that some researchers did not believe it was possible, leading to years of competing theories and intense debate. But as Darwin himself recognized, when the desires of humans meet the diversity of nature the result can indeed be astounding.

The molecular revolution of the last 2 decades has provided compelling evidence that teosinte is the progenitor of modern maize. Here, we discuss the rich genetic diversity at the source of this morphological conversion and examine how human selection has impacted this diversity. One key question concerning maize domestication remains to be resolved: was maize domestication the result of selection on a small number of loci with large effects, a large number of loci with small effects, or both? Recent genetic evidence has provided clues about the relative contributions of large-effect and small-effect loci. We discuss how future studies will help unravel the mysteries surrounding maize domestication and how this information is key to future improvements of maize.

Origins of Maize

Maize (Poaceae) is a member of the world's most successful family of agricultural crops, including wheat, rice, oats, sorghum, barley, and sugarcane. Maize belongs to the genus *Zea*, a group of annual and perennial grasses native to Mexico and Central America. The genus *Zea* includes wild taxa known collectively as teosinte (*Zea mays* ssp. *parviglumis*) and domesticated corn or maize (*Z. mays* ssp. *mays*).

For many years, relationships within the genus *Zea* were the subject of much controversy. The central difficulty in the taxonomy of maize and the identification of its closest relatives was the absence of a cob-like pistillate inflorescence, or "ear," in any other known plant. Whereas teosinte produces only 6–12 kernels in 2 interleaved rows protected by a hard outer covering (Fig. 1), modern maize boasts a cob consisting of as many as 20 rows or more with numerous exposed kernels. In fact, teosinte is so unlike maize in the structure of its ear that 19th-century botanists failed to recognize the close relationship between these plants, placing teosinte in its own genus, *Euchlaena* (2). Essentially, every new genetic technology and approach developed over the last century has been applied in an effort to resolve the question of precisely how teosinte and modern maize are related genetically.

The tremendous differences in morphology between teosinte and maize led Paul Mangelsdorf and his colleague Robert Reeves in the late 1930s to propose the Tripartite Hypothesis (3–5). This hypothesis stated that maize was domesticated from a now-extinct wild maize from South America; teosinte originated from a cross between maize and another grass, *Tripsacum*; the abundant diversity in maize was thought to be caused by "contamination" of *Tripsacum* chromosomes. This hypothesis was validated by their successful cross of maize and *Tripsacum*, although only a few, largely-sterile maize–*Tripsacum* hybrids were obtained through surgical rescue of embryos. They also analyzed backcross populations of maize–teosinte hybrids and identified 4 factors (which they interpreted as 4 *Tripsacum* chromosomal segments) responsible for the morphological differences between maize and teosinte (3–5).

For George Beadle, however, the morphological differences between maize and teosinte were not so large as to require an extinct ancestor. In his Teosinte Hypothesis, Beadle stated that maize is simply a domesticated form of teosinte (6). He believed that, through artificial selection by ancient humans, several mutations with relatively large effects could have transformed teosinte into maize. Beadle actually used Mangelsdorf and Reeves's own data against them, claiming that their 4 factors

Fig. 1. The seed spike, or ear, of teosinte (*Z. mays* ssp. *parviglumis*) consists of 2 interleaved rows of 6–12 kernels enclosed in a hard fruitcase (cupule). This female inflorescence, which differs so dramatically from that of maize, has led to much controversy and debate surrounding the origins of maize. (Photos by John Doebley.)

might just as well correspond to 4 major genes, each of which controlled a single trait that differentiated teosinte from maize. He also challenged their idea that a cross between maize and *Tripsacum*, which took Herculean efforts, would have ever occurred in nature.

Despite these profound physical differences and controversial hypotheses, various morphological, cytological, and genetic studies would eventually delineate relationships within the genus *Zea*. H.G. Wilkes laid the foundation for the current classification scheme in 1967 with the first thorough monograph on teosinte (7). This work was expanded by the rigorous evaluation of numerous traits and the discovery of many new populations by Sanchez et al. (8). In 1980, Hugh Iltis and John Doebley (9, 10) produced a system of classification that considered the probable evolutionary relationships between *Zea* taxa.

The issue was further resolved through numerous molecular and cytogenetic studies over the last century. One early indication that maize is strongly allied with teosinte came from studies of both chromosome morphology and number. Most *Zea* species and subspecies, including maize, have 10 chromosomes (11, 12), whereas most *Tripsacum* species have either 18 or 36 chromosomes (3, 4). Additionally, the cytogenetic chromosomal knobs of maize are most similar to knobs of *Z. mays* ssp. *parviglumis* and *mexicana* (5, 11, 13). Molecular genetic findings have consistently refined these relationships with data from isozymes (14), chloroplast DNA (15), and ribosomal DNA (89), all of which implicate ssp. *parviglumis* as the closest living relative of modern maize. Simple sequence repeat markers later suggested that maize was derived in a single domestication event from ssp. *parviglumis* from the Balsas River valley (17). That study revealed that ssp. *mexicana* is separated from all maize (ssp. *mays*) samples, whereas samples of ssp. *parviglumis* overlap those of maize, documenting the close relationship between ssp. *parviglumis* and maize and supporting the phylogenetic inference that the latter subspecies was the sole progenitor of maize (17).

The overall result of these analyses is that the 2 relevant subspecies of *Z. mays* (i.e., *Z. mays* ssp. *mays*, and *Z. mays* ssp. *parviglumis*) are only slightly differentiated from one another

throughout most of their genomes but have, in a very short period, evolved very different morphologies.

Maize has varied in an extraordinary and conspicuous manner.

Charles Darwin (1)

Variation: The Food of Evolution

The ability of Native Americans and modern breeders to transform a wild grass into the world's largest production grain crop is not only the product of skillful breeding, but also a tribute to the tremendous diversity of the teosinte genome. Millennia before Darwin's time, these ancient farmers first practiced what Darwin would later preach, that selection must be combined with natural variation for evolution to take place. As it turns out, teosinte is extremely diverse, with modern molecular studies measuring nucleotide diversity at silent sites in *Z. mays* ssp. *parviglumis* at ≈2–3% (18–22). This begs the question as to why *Z. mays* ssp. *parviglumis* has such high genetic diversity. In general, population genetic theory predicts that the level of selectively neutral molecular diversity is a joint function of mutation rate and effective population size, both of which would seem to be large in *Z. mays* ssp. *parviglumis*. A high rate of mutation has been documented in grasses (23), and population size for this wild grass has historically been quite large, especially for the teosintes near maize's region of origin (24).

Any 2 maize varieties differ from one another in 1.4% of their DNA (silent sites) (25). This level of nucleotide diversity is 2- to 5-fold higher than that of other domesticated grass crops and 14 times higher than that of humans. Indeed the divergence between 2 maize lines is approximately equivalent to the difference between humans and chimpanzees (26). This high level of genetic diversity results mainly from the unusually large amount of genetic diversity in its wild progenitor, *Z. mays* ssp. *parviglumis*, and the absence of a severe domestication bottleneck. Indeed, maize has apparently maintained a substantial proportion (60–70%) of the variation of its wild progenitor (25, 49), probably because humans (both ancient and modern) rely heavily on domesticated corn as a basis for subsistence, requiring thousands of plants to produce sufficient food for even small family groups (16, 27). However, this is not the case for some other domesticated crops. For example, tomato experienced a very severe genetic bottleneck as the crop was carried from the Andes to Europe, resulting in a loss of ≈95% of genetic diversity (28, 29), probably because the selection of a horticultural crop like tomato is usually done on a single plant basis with small numbers of selected plants (28).

What Were the Steps of Domestication Like?

On the surface, both dueling hypotheses (i.e., Tripartite Hypothesis and Teosinte Hypothesis) focused on the origins of corn, but at the core of the controversy was a Darwinian debate that was much more fundamental and far-reaching. In one corner were evolutionary traditionalists who held that evolution proceeds slowly over time, because of the accumulation of many small changes in numerous genes. For them, the dramatic transformation from teosinte to maize was deemed impossible in the mere 10,000 years that humans have been domesticating plants, and a more "logical" starting point for natural selection was needed. In the other corner were people like Beadle and Emerson who saw evolution as being more rapid if propelled by changes in a few significant genes. So, although teosinte and maize look strikingly different, the observed differences might be accounted for by only a few major genes, thus explaining why the 2 plants were otherwise genetically similar.

Indeed, the size of the individual evolutionary step depends strongly on the trait under consideration. As we will show below, the initial morphological changes that enabled the shift from a wild

Tian et al.

grass to a grass whose reproduction depends on humans likely involved only a few genes with large effects. Adaptation to different environments and the increase in harvestable yield, however, probably involved thousands of genes with small effects.

Recent quantitative trait loci (QTL) analyses have provided evidence supporting the notion that a few regions of the maize genome specify the key traits that distinguish maize from teosinte. Beadle conducted what could be considered the intellectual precursors of such QTL analyses. Using basic Mendelian ratios from 50,000 maize and teosinte hybrids, Beadle (30–32) recognized that as few as 5 loci might be involved in important ear and plant morphological changes. More than 20 years later, QTL mapping would validate this hypothesis, identifying 5 regions of the maize genome with large effects on basic morphology (33, 34). Two of these regions have now been characterized thoroughly.

A single major locus, *teosinte glume architecture1* (*tga1*), has been identified as a QTL controlling the formation of the tough protective covering on teosinte kernels that is mostly lacking in maize (35). The stone-like fruitcase surrounding teosinte kernels assures their unscathed passage through an animal's digestive tract, allowing seed dispersal. Because teosinte's hard glumes made it very difficult to eat, Native Americans were likely growing, harvesting, and grinding teosinte kernels themselves before the mutation leading to a softer glume came along. Thus, this mutation was probably among the first targets of selection during the domestication process. We now know that a single amino acid mutation in a transcription factor is the most likely cause of this radical change (36). And given the radical change in phenotype, it is not surprising that this mutation is not present in teosinte, which would likely be very deleterious in the wild. Given a number of assumptions, the selection intensity can be estimated at 3–4% (36). Despite this high intensity, the genomic region encompassed by this selective sweep is relatively small (the 3′ end of the gene retains substantial diversity in common with teosinte), an outcome that appears to be consistent with the maintenance of rather large population sizes and relatively unrestricted recombination throughout the domestication process.

A second locus, *teosinte branched1* (*tb1*), which dictates a difference in plant architecture (long lateral branches terminated by male tassels in teosinte vs. short lateral branches tipped by female ears in maize) has been successfully cloned (37–39). Because this locus represents a key step in maize domestication, its nucleotide diversity should be reduced when compared with neutral sites. Indeed, within the promoter region of *tb1*, maize possesses only 3% of the diversity found in teosinte (38). As is also true for *tga1*, selection does not appear to have reduced diversity throughout the entire gene. However, the low-diversity region extends 60 kb upstream regions containing some repetitive DNA but no other genes (40). Although there is some evidence for multiple functional elements in *tb1*, the major element is 60 kb upstream of the gene (41). There is also evidence for a second, distant interfering sweep at this locus (42). The timing and sequence of such character selection by early farmers is now being revealed by the fusion of molecular biology and archaeological research. Surveys of *tb1* in ancient DNA have suggested that selection at this locus occurred 4,400 years ago (43). It appears that the allele for this transformation is present in teosinte, but this possibility needs to be tested rigorously (i.e., by unraveling the full allelic series). The fitness of this locus in wild plants and whether the gene might be advantageous in particular environments is also unclear.

The large phenotypic effects of *tga1* and *tb1* undoubtedly facilitated their molecular cloning. However, how representative are these genes for the genetic basis of the domestication syndrome? With the development of a larger QTL mapping population with more power to detect QTL, Briggs et al. (44)

were able to identify more regions that contributed to the morphology of domesticated maize. In total, they detected 314 QTLs for 22 morphological traits over 2 locations. Of these, only 14 QTLs individually explained >10% of the phenotypic variation in a given trait. Most of these 14 QTLs are large-effect loci identified as essential for the transformation of teosinte to maize (45). The number of QTLs detected per trait varied substantially from 6 to 26. Interestingly, for some traits they did not detect large-effect QTL but only a number of small-effect QTLs. These results suggested that although a few genes may make the species dependent on humans for propagation, the subsequent process of genetic modification to meet human needs such as increased harvestable yield and better kernel quality or adaptation to local environments might have involved more loci with small effects, resulting in a more complex evolutionary pattern. Indeed, large-scale surveys of molecular diversity have indicated that thousands of genes might have been involved in the domestication and improvement processes. Recently, the characterization of agronomically important pathways and the dissection of complex traits have further enhanced our understanding of maize domestication.

Surveys of random markers and genes throughout the maize genome suggested that numerous genes have been targets of selection since domestication (46–49). In screens of microsatellites, ≈5% of the genome was deduced from indirect evidence to have been targeted by selection (46, 47). In their survey of 774 maize genes, Wright et al. (49) provided another estimate of the proportion of the genes under selection: ≈2–4%. If the maize genome contains 59,000 genes, Wright et al.'s estimation suggested that a minimum of 1,200 genes throughout the genome have been targets of selection during maize domestication.

Starch is the key product of maize, accounting for 73% of the kernel's total weight. The genes involved in starch synthesis are among the most important for grain production, critical to both the yield and the quality of the grain. Association analysis of 6 major candidate genes involved in starch biosynthesis (*ae1*, *bt2*, *sh1*, *sh2*, *su1*, and *wx1*) revealed that 4 of them were significantly associated with either starch concentration or composition, each of which explained <10% of phenotypic variation (50). The survey of the nucleotide diversity and selection testing at these 6 genes was striking. Four of the 6 loci exhibited evidence of selection (22). Ancient DNA analysis from maize samples unearthed in Mexico and the southwestern United States has also revealed that the *su1* alleles known to occur in modern maize were likely under selection between 1,800 and 900 years ago (43). These results suggested that Native Americans and modern breeders might have focused on improving the yield of starch and favored different amylopectin qualities.

The genetic dissection of maize flowering time and kernel composition has also argued that, despite large changes in phenotype, the alleles at the basis of these traits generally have small effects. Today, maize landraces flower from 36 days to >180 days after planting (M. Goodman, personal communication). But even at the *Vgt1* locus, one of the biggest effect genes involved in the adaptation to northern environments, has only a 1.5-day effect (51) (J. Peiffer, personal communication).

To investigate the genetic architecture of kernel oil content, the University of Illinois has conducted the world's longest controlled selection experiment; they have selected for maize with high and low kernel oil concentration for >70 generations. This selection has expanded the range of phenotypic variation ≈20-fold between high and low oil lines. However, a molecular QTL analysis suggested that >50 genes control the variation with no major genes (52).

Selection on Chromosome 10

Several large-effect QTLs for local adaptation (53–56) and domestication traits (44) have been localized to the vicinity of bin
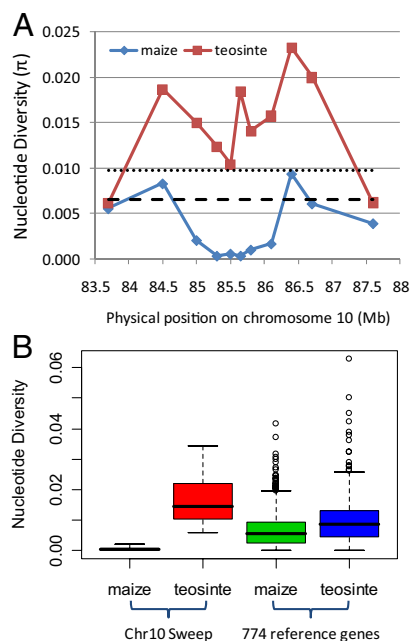
Fig. 2. Nucleotide variation of studied regions on chromosome 10. (*A*) Nucleotide diversity (π) for maize and teosinte along the investigated regions on chromosome 10. The dotted line and dash line represent the average nucleotide diversity of 774 genes (49) in teosinte and maize samples, respectively. (*B*) The comparison of nucleotide diversity (π) between chromosome 10 selective sweep and 774 reference genes (49).

10.04 on chromosome 10. These studies suggest that recent positive selection for domestication and adaptation traits may have played an important role in shaping patterns of genetic diversity in this region. We evaluated this hypothesis by resequencing a diverse panel of maize and teosinte germplasm (see *Materials and Methods*). Here, we demonstrate that this region exhibits a more extensive signal for positive selection than any other known region in the maize genome.

Initially, sequencing of candidate genes under a chromosome 10 QTL peak highlighted *ZmETR2*, a maize orthologue of the *Arabidopsis* ethylene receptor *ETR2* (57). *ZmETR2* had unusually low genetic diversity in maize relative to teosinte, suggesting possible selection at this locus. To investigate the signature of selection in this region in more detail, we sequenced 22 loci spanning ≈4 Mb in a panel of 28 diverse maize inbreds and 16 teosinte (*Z. mays* ssp. *parviglumis*) inbreds (see *Materials and Methods*). Maize exhibits severely reduced nucleotide diversity relative to teosinte across a 1.1-Mb region of chromosome 10 (Fig. 2*A*). Only 3.6% of the silent site diversity was retained in maize as compared with teosinte (Table 1). This extreme reduction of diversity suggested that functional variants within this 1.1-Mb region might have experienced recent and strong positive selection.

To delimit the region affected by the selective sweep, we performed 3 selection tests. (*i*) We compared the observed heterozygosity in maize and teosinte samples to those from simulations to determine whether the empirical pattern is significantly different from that expected under the standard neutral demographic models. We detected significant deviations from neutral expectation at loci 4–20 in maize ($P < 0.05$) (Table 1). No significant deviations from neutral expectations were observed at loci 1–3 and loci 21–23 in maize samples and all loci in teosinte samples. (*ii*) We used the HKA test (58) to examine within-species polymorphisms and between-species divergence. Under the neutral theory of molecular evolution, the amount of within-species diversity should be correlated with levels of

between-species divergence (59). We first examined the heterogeneity of the polymorphism to divergence ratio across the studied loci. We found significant heterogeneity across investigated loci in maize ($\chi^2 = 17.65$, $P = 0.016$), suggesting that these loci have experienced different evolutionary histories. Then, using 9 known unlinked neutral genes (25) as controls, we found significant departures from neutral expectation at loci 8–13 ($\chi^2 = 54.08$, $P < 0.0001$) in maize. Locus 3 is marginally significant ($\chi^2 = 12.83$, $P = 0.095$), and loci 2 and 23 are not significant ($P = 0.186$ and 0.119, respectively). We did not detect any significant departure from neutral expectations in teosinte samples in any of the tests. Moreover, we can exclude selective constraints and low mutation rates as reasons for the observed pattern because neither divergences in maize or teosinte were found to be significantly different from the genomewide average when using *Tripsacum dactyloides* as an outgroup (Table 1). (*iii*) We evaluated the probability of the observed reduction of genetic diversity in maize relative to teosinte under the neutral maize domestication bottleneck model. We simulated a population bottleneck for each studied locus by using parameters of the maize domestication bottleneck model established in Wright et al. (49) (see *Materials and Methods*). Significant deviations from expectations under a neutral domestication bottleneck were detected at loci 3–20 ($P < 0.05$), suggesting that the severe loss of genetic diversity at loci 3–20 in maize relative to teosinte cannot be explained by the maize domestication bottleneck alone. Thus, selection might have strongly shaped the genetic diversity of these loci.

Phylogenetic analysis of the investigated region revealed a star-like phylogeny within this swept region, a typical characteristic of selective sweep (60, 61). Outside of the swept regions, however, the genealogies reverted to neutral expectation, with all maize samples interspersed with teosinte samples.

The large interval (1.1 Mb) affected by this selective sweep suggested that selection was recent and strong. However, the size of this sweep could be caused by a low local recombination rate. Preliminary evidence from mapping populations suggested that the local recombination is suppressed ≈5-fold (M. McMullen, personal communication). Indeed, the nucleotide estimates of population recombination rate indicated that the recombination rate across the sweep region ($R_n = 0.0414$) was ≈3-fold lower than it was at the known selection target, *tga1* [$R_n = 0.1205$ (36)]. This result suggested that low recombination has contributed to the size of the sweep, but probably, low recombination does not fully explain the 10- to 30-fold difference in size of the selective sweeps between the chromosome 10 region and those at *tga1* and *tb1*.

We next assessed the strength of selection responsible for the chromosome 10 sweep. Kim and Stephan (62) proposed a composite likelihood ratio (CLR) test for detecting positive selection along a recombining chromosome. This method compares the likelihood of the observed pattern of nucleotide sequence variation under either a selective sweep or a standard neutral model. If the resulting LR ratio is significant, this test will provide estimates of selection strength and selection target. We applied the CLR test to a contiguous region in the *ZmETR2* region (loci 7–18), because discontiguous sequences increase the chance of false positives (63). This restriction made our results more conservative. We detected a significant LR for a selective sweep model versus a neutral model (LR = 64.7, $P < 0.001$). However, because the CLR test is under the assumption of a randomly mating population of constant size, undetected population structure or a recent bottleneck might produce a similar nucleotide variation pattern as selective sweep (64). To accommodate this weakness in the CLR test, Jensen et al. (64) proposed a goodness of fit (GOF) test to discriminate between positive selection and nonselective effects. In the GOF test, the selection scenario produced by the CLR test was used as null distribution to evaluate the significance of observed GOF value.

Table 1. Summary of sequence data of chromosome 10 regions investigated

Locus	Position, Mb	Maize							Teosinte							$\pi_{maize}/\pi_{teosinte}$, %
		N	L	S	π_{silent}	θ_{silent}	Tajima's D	k	N	L	S	π_{silent}	θ_{silent}	Tajima's D	k	
1	83.70	14	191	2	0.0056	0.0032	1.932	NA	13	198	4	0.0062	0.0065	−0.153	NA	89.7
2	84.50	27	497	4	0.0083	0.0069	0.550	0.024	15	462	9	0.0187	0.0201	−0.524	0.023	44.4
3	85.00	21	377	2	0.0021	0.0061	−1.514	0.027	8	467	11	0.0150	0.0138	−0.958	0.028	14.0*
4	85.30	26	413	1	0.0004**	0.0006**	−0.714	NA	12	400	19	0.0124	0.0157	−0.930	NA	3.2**
5	85.50	27	474	0	0.0000**	0.0000**	NA	NA	11	475	25	0.0152	0.0180	−0.718	NA	0.0**
6	85.50	28	334	1	0.0006**	0.0008*	−0.363	NA	12	336	15	0.0104	0.0148	−1.268	NA	5.8**
7	85.65	27	405	0	0.0000**	0.0000**	NA	NA	9	429	16	0.0157	0.0146	0.385	NA	0.0**
8	85.65	28	446	1	0.0002**	0.0006**	−1.151	0.052	11	465	12	0.0061	0.0066	−0.758	0.054	3.3**
9	85.65	25	559	1	0.0000**	0.0000**	−1.151	0.041	10	683	26	0.0219	0.0230	−0.283	0.05	0.0**
10	85.65	28	626	1	0.0016**	0.0009**	1.213	0.026	12	627	11	0.0132	0.0139	−0.212	0.027	12.1*
11	85.65	27	582	1	0.0000**	0.0000**	−0.338	0.009	12	552	5	0.0142	0.0134	0.227	0.003	0.0**
12	85.65	27	716	0	0.0000**	0.0000**	NA	0.021	8	714	7	0.0086	0.0100	−0.667	0.020	0.0**
13	85.65	28	561	0	0.0000**	0.0000**	NA	0.057	11	515	14	0.0089	0.0093	−0.183	0.057	0.0**
14	85.65	27	787	1	0.0002**	0.0003**	−0.728	NA	14	751	20	0.0068	0.0084	−0.784	NA	2.9**
15	85.65	27	610	1	0.0003**	0.0004**	−0.338	NA	11	570	48	0.0290	0.0300	0.045	NA	1.0**
16	85.65	26	466	3	0.0009**	0.0017*	−1.090	NA	8	558	51	0.0344	0.0353	0.035	NA	2.6**
17	85.65	28	660	3	0.0006**	0.0012**	−1.165	NA	10	652	64	0.0338	0.0352	−0.200	NA	1.8**
18	85.65	26	501	3	0.0006**	0.0016*	−1.513	NA	11	557	48	0.0296	0.0307	−0.165	NA	2.0**
19	85.80	27	720	9	0.0010**	0.0036	−2.275**	NA	11	750	29	0.0142	0.0129	−0.073	NA	7.0**
20	86.10	26	421	3	0.0017**	0.0019*	−0.216	NA	14	461	28	0.0158	0.0191	−0.743	NA	10.8**
21	86.40	26	595	21	0.0094	0.0161	−1.496	NA	13	586	28	0.0233	0.0261	−0.421	NA	40.3
22	86.70	27	487	12	0.0061	0.0089	−0.846	NA	15	467	22	0.0200	0.0206	−0.241	NA	30.5*
23	87.60	25	670	5	0.0039	0.0023	1.967	0.025	11	666	12	0.0063	0.0068	−0.411	0.026	61.9

Locus 1, pza00048 obtained from Wright et al. (49); 2, pzb03525; 3, pzb03526; 4, pzb03527; 5, pzb03528; 6, pzb03529; 7–18, pzb03530–pzb03532, pzb02831, pzb03533-pzb03540; 19, pzb03541; 20, pzb03542; 21, pzb03543; 22, pzb02753; 23, pzb02851. The relative physical locations of investigated regions are based on the Maize Agarose FPC Map (www.genome.arizona.edu/fpc/maize). N, number of sequence; L, the length of a given locus; S, number of segregating sites; π and θ were estimated for noncoding and synonymous sites; K, divergence between maize or teosinte and *Tripsacum dactyloides*. NA, not available.
*Significant at the 0.05 level.
**Significant at the 0.01 level.

We detected a nonsignificant GOF value ($\wedge_{GOF} = 52.37$; $P = 0.661$), suggesting that our rejection of the neutral model in CLR test is not caused by population structure or demographic forces. In other words, positive selection rather than demography is the likely cause of the pattern observed. Furthermore, the estimated selection strength parameter ($2Ns = 2,2187.8$) is far greater than the value for *tga1* ($2Ns = 9,232$) (36). Assuming an effective population size for maize of 100,000 (36), the selection coefficients for *tga1* and the chromosome 10 selective sweep are 0.046 and 0.111, respectively. Because we used a more conservative parameter of θ, the selection coefficient of the selective sweep on chromosome 10 is at least 2.4 times larger than that of *tga1*. Strong selection strength surely contributed to the size of the sweep.

Although this large sweep region was not detected by a 774-gene survey by Wright et al. (49), large sweeps like this have been found in other situations. Among American maize varieties in the 1920s, there was also very strong selection for yellow color, which produced a large sweep around the *y1* locus in maize breeding lines (65, 66). A similar pattern was observed at the *waxy* locus in rice (67) and the *Sod* locus in *Drosophila melanogaster* (68). The challenge for the future is to reconcile how these extremely different patterns of evolution and selection have occurred.

The Future of the Revolution

It was 150 years ago that Darwin so skillfully used domesticated plants and animals to help argue for evolution and natural selection. Over the last 100 years, maize has been a prime example for studying evolution, and tremendous strides have been made in understanding its origins, the genetics of the evolution, the strength of selection, and the archaeological context. For understanding selection, the patterns are tremendously varied. We have examples of nearly every type of selection footprint: on standing variation and on novel mutations, a few large-effect genes and large numbers of selected genes with apparently small phenotypic effects, small single selective sweeps, and a few massive sweeps. Because we are in the midst of a genomics revolution, tremendous opportunities exist to advance our understanding of molecular processes. Over the next 2 decades we should be able to identify more genes involved in domestication, pinpoint the allelic variants favored through time, and evaluate successful and failed genetic alterations through time and space. Top-down and bottom-up approaches (69) are complementary and can be combined to improve our understanding of the domestication process.

The bottom-up approaches are being supercharged by next-generation sequencing, which is providing tremendous opportunities for understanding the regions of selection across the maize genome. With the completion of maize genome sequencing, and the first-generation resequencing to produce a high-resolution maize HapMap, we should soon be able to screen the entire genome for selection and identify a basic set of genes that have been targeted by selection during the domestication of maize. Indeed, we might start to evaluate the relative importance of all kinds of selection patterns. However, such genomic analyses alone will not provide precise information on when, where, and why these regions were targets of selection.

How did varieties of maize adapt to diverse environments throughout the globe? Through the exceptional efforts of maize germplasm curators over the last century, well over 20,000 landraces of maize have been collected throughout the Americas. By combining a global sample of landraces (70) with whole genome sequencing, adaptation can be evaluated. Additionally,

by relating presumptive adaptations to the increasingly rich Geographic Information System databases on climate and soils, the polymorphisms involved with environmental differentiation may be identified. Although similar molecular studies have been conducted to understand human differentiation around the globe (71), these targets of differentiation in maize can be studied experimentally and, most importantly, can be applied to adapting future maize varieties to the world's rapidly changing environments.

What is the timing and tempo of these selection events? The tempo of selection was important to Darwin and is still a central issue today. With molecular data, we can estimate the intensity of selection and the time since a selective sweep. However, each assessment requires a number of assumptions that include modeling population size, historical recombination rates, migration, and mutation rates. Although there have been some tremendous strides in such modeling, domesticated crops provide a great opportunity to empirically test particular assumptions. Millions of archaeological botanical samples are available for DNA analysis that can provide hard data on the progress of selection at particular places and times (72, 73). Small-scale studies of this style have been conducted in maize (43, 74), but the future of sequencing whole genomes from well-preserved maize paleobotanic materials is very exciting.

The reason why particular genes have been under selection is a much more difficult question, but resolving it has important implications for future crop development. In the case of *tb1* and t*ga1*, we know much about why these loci were selection targets; but in the case of the chromosome 10 region the reasons are currently much less clear. Did selection target only 1 gene and 1 trait in this region or multiple genes and traits? We will not know until mapping identifies the causative nucleotides. Although, with these approaches, it can take years to find a single gene, the maize community is now assembling an unrivaled set of tools for forward trait dissection that will greatly accelerate the process. Altogether, ≈15,500 maize and teosinte genetic stocks have been constructed, which ultimately may permit the dissection of virtually any trait (www.panzea.org). With next-generation sequencing of key founders of this germplasm and community-wide efforts to phenotype a wide range of traits, the top-down approach will likely accelerate rapidly. We expect to make regular connections between top-down and bottom-up approaches.

Maize is at the crossroads of 2 great legacies. Native Americans and nature have worked to produce a species with tremendous natural variation and selective potential that has adapted to numerous environments. The Darwinian intellectual revolution, enabled by modern technology, allows us to understand how maize arrived at its current position and provides the tools to mold maize ever more efficiently for new societal needs with directed evolution. In fact, we are continuing to follow the steps of early Native Americans who transformed teosinte into maize millennia ago. Through allele mining in existing germplasm, beneficial alleles can be discovered and potentially applied to practical breeding. Wild relatives can also be tapped to recover superior alleles that have been lost during domestication and improvement processes. Guided by lessons from past domestication, we are practicing selection magic to pyramid useful genes to produce best varieties.

Materials and Methods

Plant Materials and DNA Sequencing. We sampled DNA sequence diversity in a panel of 28 diverse maize inbreds and 16 teosinte (*Z. mays* ssp. *parviglumis*) inbreds. The panel was selected to maximize the genetic diversity of maize (75, 76, 77) and represents a wide geographical distribution of wild teosinte germplasm (www.panzea.org). A total of 28 maize inbred lines are 26 founders of Nested Association Mapping (NAM) (75) population and 2 other inbred lines, Mo17 and W22 R-r:std. Sixteen teosinte inbred lines (TIL01-TIL12 and TIL14-TIL17) were kindly provided by John Doebley (University of Wisconsin, Madison). A *Tripsacum dactyloides* sample (MIA34597) was used as an outgroup to estimate divergence. A total of 23 loci were surveyed to identify the physical boundary of the selective sweep region based on the maize FPC map (www.genome.arizona.edu/fpc/maize). Sequencing reactions were performed on PCR products in both directions with BigDye v3.1 on an Applied Biosystem 3730 automated sequencer. Base calling, quality checks, and sequence assembly were conducted with PHRED and PHRAP (78). Multiple sequence alignments were made by using Biolign (http://en.bio-soft.net/dna/BioLign.html) and manually edited if necessary.

Data Analysis. The number of segregating sites (*S*), the nucleotide diversity θ (79) and π (80) at silent sites, the divergences in maize and in teosinte from *Tripsacum*, and Tajima's *D* statistic (81) were estimated by using DNAsp 4.10 (82). Insertions and deletions were not included in the analysis. We used the multilocus Hudson-Kreitman-Aguade (HKA) test (58) to test the ratios of DNA sequence polymorphisms to divergence across loci using the *Tripsacum dactyloides* sequence as an outgroup. We used Hudson's ms program (83) to do 10,000 coalescent simulations to estimate the probability of observing a given level of genetic diversity under a standard neutral model with the conservative assumption of no recombination (84). The expected heterozygosity implemented in the simulation was $\theta = 0.0064$ and 0.0112 in maize and teosinte, respectively, estimated from 774 reference genes (49). Coalescent simulations that incorporated the domestication bottleneck (19, 85) were performed for each studied locus with the ms program. All parameters in the model were assigned to the established values (49). Based on a survey of 774 genes, the best fit of the severity of maize domestication bottleneck (*k*), the ratio of population size during bottleneck (*Nb*) to the duration of bottleneck (*d*), was 2.45 (49). The population mutation parameter θ (79) and population recombination parameter *4Nc* (86) were estimated from the teosinte data. Using the neutral domestication bottleneck as the null distribution, we evaluated the probability of the observed loss of genetic diversity in maize relative to teosinte based on 10,000 coalescent simulations. The CLR test proposed by Kim and Stephan (62) was used to test the hitchhiking effect and estimate the selection coefficient. We focused this analysis on the *ZmETR2* region (loci 7–18) containing ≈7 kb of contiguous sequence. Ancestral and derived alleles at polymorphic sites were identified by comparing to the *Tripsacum* sequence. If the derived state of a segregating site could not be determined because of unavailable *Tripsacum* sequence, we assumed the base with the higher frequency to be ancestral. This assumption is conservative and has little effect in detecting selection (62, 87). In those loci with a missing state for particular lines, we assumed the segregating sites at these missing sequences had the ancestral state, which is a conservative assumption as shown by the study of Orengo and Aguadé (88). We did not provide a selection target estimation for 2 reasons: (*i*) a partially-sequenced region will give a less reliable estimate of the selection target (63, 64); (*ii*) this selective sweep affected so many regions that estimating the selection target based on a single region is not meaningful. The basic analysis strategy of the CLR test is the same as that described by Wang et al. (36) with minor modifications. Instead of estimating θ from local teosinte data as Wang et al. (36) did for t*ga1*, we used a more conservative estimate of $\theta = 0.0064$, estimated from a genomewide value (49) as the expected nucleotide diversity in maize. The scaled per-nucleotide recombination parameter $R_n = 0.0414$ (86) is the length-weighted mean of R_n across the *ZmETR2* region (loci 7–18) estimated from teosinte data. The significance of the resulting likelihood ratio was evaluated by 1,000 simulations of neutral datasets. The GOF test (64) was further used to distinguish between selective sweep and demographic forces. The significance of the GOF value for the observed data were evaluated by 1,000 simulations under the selection scenario produced by the above CLR test.

1. Darwin CR (1868) *The Variation of Animals and Plants Under Domestication* (Orange Judd & Co., New York), 1st Ed.
2. Doebley J (1990) Molecular systematics of *Zea* (Gramineae). *Maydica* 35:143–150.
3. Mangelsdorf PC, Reeves RG (1938) The origin of maize. *Proc Natl Acad Sci USA* 24:303–312.
4. Mangelsdorf PC, Reeves RG (1939) *The Origin of Indian Corn and Its Relatives* (Texas Agricultural Experiment Station, College Station), Bulletin 574.
5. Mangelsdorf PC (1974) *Corn: Its Origin Evolution and Improvement* (Harvard Univ Press, Cambridge, MA).
6. Beadle GW (1939) Teosinte and the origin of maize. *J Hered* 30:245–247.

7. Wilkes HG (1967) *Teosinte: The Closest Relative of Maize* (Harvard Univ Cambridge, MA).

8. Sanchez G, et al. (1998) *Distribucion y Caracterizacion del Teocinte* (Instituto Nacional de Investigaciones Forestales, Agricolas y Pecuarias, Guadalajara, Mexico).

9. Iltis HH, Doebley JF (1980) Taxonomy of *Zea* (Gramineae). II. Subspecific categories in the *Zea mays* complex and generic synopsis. *Am J Bot* 67:994–1004.

10. Doebley JF, Iltis HH (1980) Taxonomy of *Zea* (Gramineae) I. A subgeneric classification with key to taxa. *Am J Bot* 67:982–993.

11. Kato YTA (1976) Cytological studies of maize (*Zea mays* L.) and teosinte (*Zea mexicana* Schrader Kuntze) in relation to their origin and evolution. *Mass Agric Exp Sta Bull* 635:1–185.

12. Kato YTA, Lopez RA (1990) Chromosome knobs of the perennial teosintes. *Maydica* 35:125–141.

13. McClintock B, Kato YTA, Blumenschein A (1981) *Chromosome Constitution of Races of Maize* (Colegio de Postgraduados, Chapingo, Mexico).

14. Doebley JF, Goodman MM, Stuber CW (1984) Isoenzymatic variation in *Zea* (Gramineae). *Syst Bot* 9:203–218.

15. Doebley J, Renfroe W, Blanton A (1987) Restriction site variation in the *Zea* chloroplast genome. *Genetics* 117:139–147.

16. Buckler ES, IV, Thornsberry JM, Kresovich S (2001) Molecular diversity, structure, and domestication of grasses. *Genet Res* 77:213–218.

17. Matsuoka Y, et al. (2002) A single domestication for maize shown by multilocus microsatellite genotyping. *Proc Natl Acad Sci USA* 99:6080–6084.

18. White SE, Doebley JF (1999) The molecular evolution of *terminal ear1*, a regulatory gene in the genus *Zea*. *Genetics* 153:1455–1462.

19. Eyre-Walker A, Gaut RL, Hilton H, Feldman DL, Gaut BS (1998) Investigation of the bottleneck leading to the domestication of maize. *Proc Natl Acad Sci USA* 95:4441–4446.

20. Hilton H, Gaut BS (1998) Speciation and domestication in maize and its wild relatives: Evidence from the globulin-1 gene. *Genetics* 150:863–872.

21. Goloubinoff P, Pääbo S, Wilson AC (1993) Evolution of maize inferred from sequence diversity of an *Adh2* gene segment from archaeological specimens. *Proc Natl Acad Sci USA* 90:1997–2001.

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

23. Gaut BS, Morton BR, McCaig BC, Clegg MT (1996) Substitution rate comparisons between grasses and palms: Synonymous rate differences at the nuclear gene *Adh* parallel rate differences at the plastid gene *rbcL*. *Proc Natl Acad Sci USA* 93:10274–10279.

24. 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.

25. Tenaillon MI, et al. (2001) Patterns of DNA sequence polymorphism along chromosome 1 of maize (*Zea mays* ssp. *mays* L). *Proc Natl Acad Sci USA* 98:9161–9166.

26. Chen FC, Li WH (2001) Genomic divergences between humans and other hominoids and the effective population size of the common ancestor of humans and chimpanzees. *Am J Hum Genet* 68:444–456.

27. Hillman G, Davies MS (1990) Domestication rates in wild-type wheats and barley under primitive cultivation. *Biol J Linn Soc* 39:39–78.

28. Bai Y, Lindhout P (2007) Domestication and breeding of tomatoes: What have we gained and what can we gain in the future? *Ann Bot* 100:1085–1094.

29. Miller JC, Tanksley SD (1990) RFLP analysis of phylogenetic relationships and genetic variation in the genus *Lycopersicon*. *Theor Appl Genet* 80:437–448.

30. Beadle GW (1980) The ancestry of corn. *Sci Am* 242:112–119.

31. Beadle GW (1977) The origin of *Zea mays*. *Origins of Agriculture*, ed Reed CA (Mouton Press, The Hague), pp 615–635.

32. Beadle GW (1972) The mystery of maize. *Field Museum Natural History Bull* 43:2–11.

33. Doebley J, Stec A, Wendel J, Edwards M (1990) Genetic and morphological analysis of a maize-teosinte F$_2$ population: Implications for the origin of maize. *Proc Natl Acad Sci USA* 87:9888–9892.

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

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

36. Wang H, et al. (2005) The origin of the naked grains of maize. *Nature* 436:714–719.

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

38. Wang RL, Stec A, Hey J, Lukens L, Doebley J (1999) The limits of selection during maize domestication. *Nature* 398:236–239.

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

40. Clark RM, Linton E, Messing J, Doebley JF (2004) Pattern of diversity in the genomic region near the maize domestication gene *tb1*. *Proc Natl Acad Sci USA* 101:700–707.

41. Clark RM, Wagler TN, Quijada P, Doebley J (2006) A distant upstream enhancer at the maize domestication gene *tb1* has pleiotropic effects on plant and inflorescence architecture. *Nat Genet* 38:594–597.

42. Camus-Kulandaivelu L, et al. (2008) Patterns of molecular evolution associated with two selective sweeps in the *Tb1-Dwarf8* region in maize. *Genetics* 180:1107–1121.

43. Jaenicke V, et al. (2003) Early allelic selection in maize as revealed by ancient DNA. *Science* 302:1206–1208.

44. Briggs WH, McMullen MD, Gaut BS, Doebley J (2007) Linkage mapping of domestication loci in a large maize-teosinte backcross resource. *Genetics* 177:1915–1928.

45. Doebley J, Stec A (1993) Inheritance of the morphological differences between maize and teosinte: Comparison of results for two F$_2$ populations. *Genetics* 134:559–570.

46. Vigouroux Y, et al. (2002) Identifying genes of agronomic importance in maize by screening microsatellites for evidence of selection during domestication. *Proc Natl Acad Sci USA* 99:9650–9655.

47. Vigouroux Y, et al. (2005) An analysis of genetic diversity across the maize genome using microsatellites. *Genetics* 169:1617–1630.

48. Yamasaki M, et al. (2005) A large-scale screen for artificial selection in maize identifies candidate agronomic loci for domestication and crop improvement. *Plant Cell* 17:2859–2872.

49. Wright SI, et al. (2005) The effects of artificial selection on the maize genome. *Science* 308:1310–1314.

50. Wilson LM, et al. (2004) Dissection of maize kernel composition and starch production by candidate gene association. *Plant Cell* 16:2719–2733.

51. Salvi S, et al. (2007) Conserved noncoding genomic sequences associated with a flowering-time quantitative trait locus in maize. *Proc Natl Acad Sci USA* 104:11376–11381.

52. Laurie CC, et al. (2004) The genetic architecture of response to long-term artificial selection for oil concentration in the maize kernel. *Genetics* 168:2141–2155.

53. Bouchez A, Hospital F, Causse M, Gallais A, Charcosset A (2002) Marker-assisted introgression of favorable alleles at quantitative trait loci between maize elite lines. *Genetics* 162:1945–1959.

54. Ribaut JM, Hoisington DA, Deutsch JA, Jiang C, Gonzalez-de-Leon D (1996) Identification of quantitative trait loci under drought conditions in tropical maize. 1. Flowering parameters and the anthesis-silking interval. *Theor Appl Genet* 92:905–914.

55. Wang CL, et al. (2008) Genetic analysis of photoperiod sensitivity in a tropical by temperate maize recombinant inbred population using molecular markers. *Theor Appl Genet* 117:1129–1139.

56. Mano Y, Omori F, Kindiger B, Takahashi H (2008) A linkage map of maize x teosinte Zea luxurians and identification of QTLs controlling root aerenchyma formation. *Mol Breed* 21:327–337.

57. Sakai H, et al. (1998) ETR2 is an *ETR1*-like gene involved in ethylene signaling in *Arabidopsis*. *Proc Natl Acad Sci USA* 95:5812–5817.

58. Hudson RR, Kreitman M, Aguade M (1987) A test of neutral molecular evolution based on nucleotide data. *Genetics* 116:153–159.

59. Kimura M (1983) *The Neutral Theory of Molecular Evolution* (Cambridge Univ Press, Cambridge, UK).

60. Meiklejohn CD, Kim Y, Hartl DL, Parsch J (2004) Identification of a locus under complex positive selection in *Drosophila simulans* by haplotype mapping and composite-likelihood estimation. *Genetics* 168:265–279.

61. Kaplan NL, Hudson RR, Langley CH (1989) The hitchhiking effect revisited. *Genetics* 123:887–899.

62. Kim Y, Stephan W (2002) Detecting a local signature of genetic hitchhiking along a recombining chromosome. *Genetics* 160:765–777.

63. Pool JE, DuMont VB, Mueller JL, Aquadro CF (2006) A scan of molecular variation leads to the narrow localization of a selective sweep affecting both afrotropical and cosmopolitan populations of *Drosophila melanogaster*. *Genetics* 172:1093–1105.

64. Jensen JD, Kim Y, DuMont VB, Aquadro CF, Bustamante CD (2005) Distinguishing between selective sweeps and demography using DNA polymorphism data. *Genetics* 170:1401–1410.

65. Palaisa KA, Morgante M, Williams M, Rafalski A (2003) Contrasting effects of selection on sequence diversity and linkage disequilibrium at two phytoene synthase loci. *Plant Cell* 15:1795–1806.

66. Palaisa K, Morgante M, Tingey S, Rafalski A (2004) Long-range patterns of diversity and linkage disequilibrium surrounding the maize *Y1* gene are indicative of an asymmetric selective sweep. *Proc Natl Acad Sci USA* 101:9885–9890.

67. Olsen KM, et al. (2006) Selection under domestication: Evidence for a sweep in the rice waxy genomic region. *Genetics* 173:975–983.

68. Saez AG, Tatarenkov A, Barrio E, Becerra NH, Ayala FJ (2003) Patterns of DNA sequence polymorphism at *Sod* vicinities in *Drosophila melanogaster*: Unraveling the footprint of a recent selective sweep. *Proc Natl Acad Sci USA* 100:1793–1798.

69. Ross-Ibarra J, Morrell PL, Gaut BS (2007) Plant domestication, a unique opportunity to identify the genetic basis of adaptation. *Proc Natl Acad Sci USA* 104:8641–8648.

70. Vigouroux Y, et al. (2008) Population structure and genetic diversity of New World maize races assessed by DNA microsatellites. *Am J Bot* 95:1240–1253.

71. Rosenberg NA, et al. (2002) Genetic structure of human populations. *Science* 298:2381–2385.

72. Piperno DR, et al. (2007) Late Pleistocene and Holocene environmental history of the Iguala Valley, Central Balsas Watershed of Mexico. *Proc Natl Acad Sci USA* 104:11874–11881.

73. Pohl ME, Piperno DR, Pope KO, Jones JG (2007) Microfossil evidence for pre-Columbian maize dispersals in the neotropics from San Andres, Tabasco, Mexico. *Proc Natl Acad Sci USA* 104:6870–6875.

74. Lia VV, et al. (2007) Microsatellite typing of ancient maize: Insights into the history of agriculture in southern South America. *Proc Biol Sci* 274:545–554.

75. Yu J, Holland JB, McMullen MD, Buckler ES (2008) Genetic design and statistical power of nested association mapping in maize. *Genetics* 178:539–551.

76. Liu KJ, et al. (2003) Genetic structure and diversity among maize inbred lines as inferred from DNA microsatellites. *Genetics* 165:2117–2128.

77. Flint-Garcia SA, et al. (2005) Maize association population: A high-resolution platform for quantitative trait locus dissection. *Plant J* 44:1054–1064.

78. Ewing B, Hillier L, Wendl MC, Green P (1998) Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res* 8:175–185.

Tian et al.

79. Watterson GA (1975) On the number of segregating sites in genetical models without recombination. *Theor Popul Biol* 7:256–276.

80. Tajima F (1983) Evolutionary relationship of DNA sequences in finite populations. *Genetics* 105:437–460.

81. Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123:585–595.

82. Rozas J, Sanchez-DelBarrio JC, Messeguer X, Rozas R (2003) DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* 19:2496–2497.

83. Hudson RR (2002) Generating samples under a Wright–Fisher neutral model of genetic variation. *Bioinformatics* 18:337–338.

84. Hudson RR (1990) Gene genealogies and the coalescent process. *Oxford Surv Evol Biol* 7:1–42.

85. Tenaillon MI, U'Ren J, Tenaillon O, Gaut BS (2004) Selection versus demography: A multilocus investigation of the domestication process in maize. *Mol Biol Evol* 21:1214–1225.

86. Hudson RR (1987) Estimating the recombination parameter of a finite population model without selection. *Genet Res* 50:245–250.

87. DuMont VB, Aquadro CF (2005) Multiple signatures of positive selection downstream of notch on the X chromosome in *Drosophila melanogaster*. *Genetics* 171:639–653.

88. Orengo DJ, Aguade M (2007) Genome scans of variation and adaptive change: Extended analysis of a candidate locus close to the phantom gene region in *Drosophila melanogaster*. *Mol Biol Evol* 24:1122–1129.

89. Buckler ES, Holtsford TP (1996) *Zea* systematics: Ribosomal ITS evidence. *Mol Biol Evol* 13:623–632.