Supporting Information

Whipple et al. 10.1073/pnas.1102819108

SI Materials and Methods

Plant Materials and Sequence Data for Nucleotide Diversity Analysis. Six regions of *gt1* were amplified (Fig. S1) in up to 24 maize inbreds, 16 maize landrace haploids, and 16 teosinte inbreds (Table S1). The 24 maize inbreds represent much of the genetic diversity among important public lines currently available for breeding. The geographically diverse sample of maize landraces represents the genetic diversity present before modern day breeding efforts (1). The teosinte inbreds are also a geographically diverse sample, encompassing the entire natural distribution of *Z. mays* ssp. *parviglumis*. For five of the gene regions, at least one of two *T. dactyloides* accessions, which belong to the sister genus of *Zea*, were sequenced and used as outgroup individuals in the selection analyses (Table S1). For the remaining region, for which the *T. dactyloide*s sequence was unavailable, two *Zea diploperennis* alleles from different accessions were isolated and used as outgroup individuals.

We directly sequenced PCR products from the haploid plants of the maize landraces, maize inbreds, and teosinte individuals using a standard protocol (Applied Biosystems). Our DNA sources for *Z. diploperennis* and *T. dactyloides* individuals were potentially heterozygous; therefore, we cloned PCR products from these sources into TOPO-TA vectors (pCR 2.1-TOPO kit; Invitrogen) and, afterward, sequenced multiple clones to identify a single allele and correct errors introduced during PCR. The forward and reverse DNA sequences were assembled for each individual using Sequencher software (Gene Codes). Individual sequences from the maize inbreds, maize landraces, teosinte, and outgroup individuals were then manually aligned using BioEdit software (2).

Tests for Neutrality. Molecular population genetics statistics were estimated separately for the maize inbreds, maize landraces, and teosinte individuals using DnaSP (3). Nucleotide polymorphism (θ) (4) and nucleotide diversity (π) (5) were calculated on the basis of all sites. The Hudson–Kreitman–Aguadé (HKA) test (6) for neutrality and Tajima's *D*-statistics (7) were also generated using DnaSP. For the five gene regions in which a *T. dactyloides* outgroup individual was amplified, 11 neutral loci (*adh1*, *an1*, *asg75*, *bz2*, *csu381*, *csu1132*, *csu1138*, *csu1171*, *fus6*, *glb1*, and *umc128*) (8) were used for the maize landrace HKA tests. Five of these genes (*adh1*, *bz2*, *csu1138*, *csu1171*, and *glb1*) were used in the corresponding HKA tests involving maize inbreds and teosinte individuals (1). A set of four neutral genes (*adh1*, *adh2*, *glb1*, and *te1*) (1, 9) was used when conducting the maize landrace HKA neutrality test for the gene region in which *Z. diploperennis* outgroup individuals were amplified. Two of these genes (*adh1* and *glb1*) were used in the corresponding HKA tests involving the maize inbreds and teosinte individuals. Additional HKA tests were performed separately for the maize landraces and teosinte individuals for those gene regions with a *T. dactyloides* outgroup. These tests used a different set of 15 control genes (DX414418, DX414430, DX414440, DX414417, DX414419, DX414425, DX414431, DX414433, DX414434, D414435, DX414437, DX414442, DX414443, DX414448, and DX414449) that previous analysis had indicated were neutral (10). For each HKA test, an overall χ^2 value was calculated by taking the sum of the individual χ^2 values calculated using each neutral locus. These overall χ^2 values were then used to obtain overall *P* values.

Coalescence Simulation Analysis for Selection. A coalescence simulation-based test (11) was used to determine if each of the six gene regions was a potential target of selection during domestication. We used a modified version of the standard coalescence procedure (6) that incorporated the domestication bottleneck as previously described (8). All parameters in the model were assigned to previously established values (1, 8). A value of 1.8 was used to estimate the severity of the bottleneck (N_b/d). For each of the six gene regions, 10,000 simulations were conducted. The *P* value was calculated as the number of simulations that produced an *S* value equal to or less than that observed in the maize landraces. A gene region was considered to be a potential target of selection during domestication if S_{maize} was <97.5% of the S_{simul} values.

Results

We investigated whether a signature of past selection was detected in *gt1*, as would be expected if *gt1* had undergone selection during domestication or improvement. First, we estimated the ratio of nucleotide diversity (π) in maize inbreds and in maize landraces compared with teosinte (Fig. S2 and Table S2). For the three gene regions with available data, the nucleotide diversity ratios between the maize inbreds and teosinte individuals were lower than the average (0.74) from a random sample of genes across the genome (12) but higher than the average (~0.30) previously reported for genes under selection (13). The nucleotide diversity ratios between the maize landraces and teosinte individuals calculated for five of the six gene regions were similar to previous estimates (60–80%) for neutral genes (1). In the last gene region, downstream of the 3′ UTR, maize only retained 9% of the diversity found in *Z. mays* ssp. *parviglumis* teosinte. This value is only slightly higher than those observed in genes like *tga1* (5%) and upstream of *tb1* (<1%), genes that have previously been identified as undergoing selection during domestication (14, 15).

Second, we calculated Tajima's *D*-statistics to test whether there was an excess of rare mutations in any of the gene regions, which is expected under positive selection (Table S3). None of the Tajima's *D*-statistics were significant for either the maize inbreds or maize landraces. This could be attributable to the limitations of this test, which does not account for demographic history, such as population expansion or reduction, both of which have occurred in maize.

Third, we conducted several HKA neutrality tests for each of the gene regions to determine whether the ratio of diversity in our set of samples (maize inbreds, maize landraces, and teosinte individuals) compared with that of an outgroup (*T. dactyloides* or *Z. diploperennis*) was significantly different from that observed in neutral genes (Table S4). The gene region downstream of the 3′ UTR was the only region with significant maize HKA test results. The corresponding HKA test result involving teosinte was not significant, suggesting that this region underwent selection during domestication.

Lastly, we conducted a coalescence simulation-based test. The coalescence simulation test determines whether the loss of diversity in maize is greater than would be expected attributable to the domestication bottleneck alone. If there is more loss of diversity in maize than expected, it is attributed to selection during domestication. Once again, the gene region downstream of the 3′ UTR was the only gene region that the coalescence simulation test identified as being under selection (Table S4), giving further evidence that this region of the gene was under selection during domestication.

1. 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.
2. Hall T (1990) BioEdit: A user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp Ser* 41:95–98.
3. Rozas J, Sánchez-DelBarrio JC, Messeguer X, Rozas R (2003) DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* 19:2496–2497.
4. Watterson GA (1975) On the number of segregating sites in genetical models without recombination. *Theor Popul Biol* 7:256–276.
5. Tajima F (1983) Evolutionary relationship of DNA sequences in finite populations. *Genetics* 105:437–460.
6. Hudson RR, Kreitman M, Aguadé M (1987) A test of neutral molecular evolution based on nucleotide data. *Genetics* 116:153–159.
7. Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123:585–595.
8. 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.
9. White SE, Doebley JF (1999) The molecular evolution of terminal ear1, a regulatory gene in the genus Zea. *Genetics* 153:1455–1462.
10. Zhao QA, et al. (2008) The role of regulatory genes during maize domestication: Evidence from nucleotide polymorphism and gene expression. *Genetics* 178:2133–2143.
11. Innan H, Kim Y (2004) Pattern of polymorphism after strong artificial selection in a domestication event. *Proc Natl Acad Sci USA* 101:10667–10672.
12. Wright SI, et al. (2005) The effects of artificial selection on the maize genome. *Science* 308:1310–1314.
13. Whitt SR, Wilson LM, Tenaillon MI, Gaut BS, Buckler ES 4th (2002) Genetic diversity and selection in the maize starch pathway. *Proc Natl Acad Sci USA* 99:12959–12962.
14. 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.
15. Wang H, et al. (2005) The origin of the naked grains of maize. *Nature* 436:714–719.

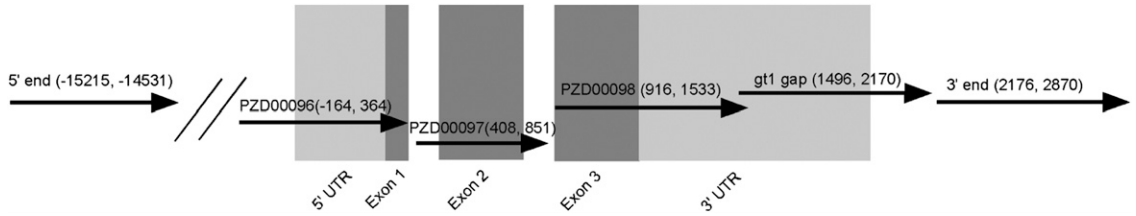

Fig. S1. *gt1* annotated with sequenced amplicons and predicted genic regions. The beginning and end bases of the amplicons are listed in parentheses, where one is defined as the first base of the predicted 5′ UTR. All positions were determined by blasting to clone AC195802.3: 1-107710 provided by the Maize Genome Browser. The clone spans all six amplicons. The shaded areas are predicted gene regions, the darkly shaded regions are predicted exons, and the lighter shaded regions represent the 3′ and 5′ UTRs.

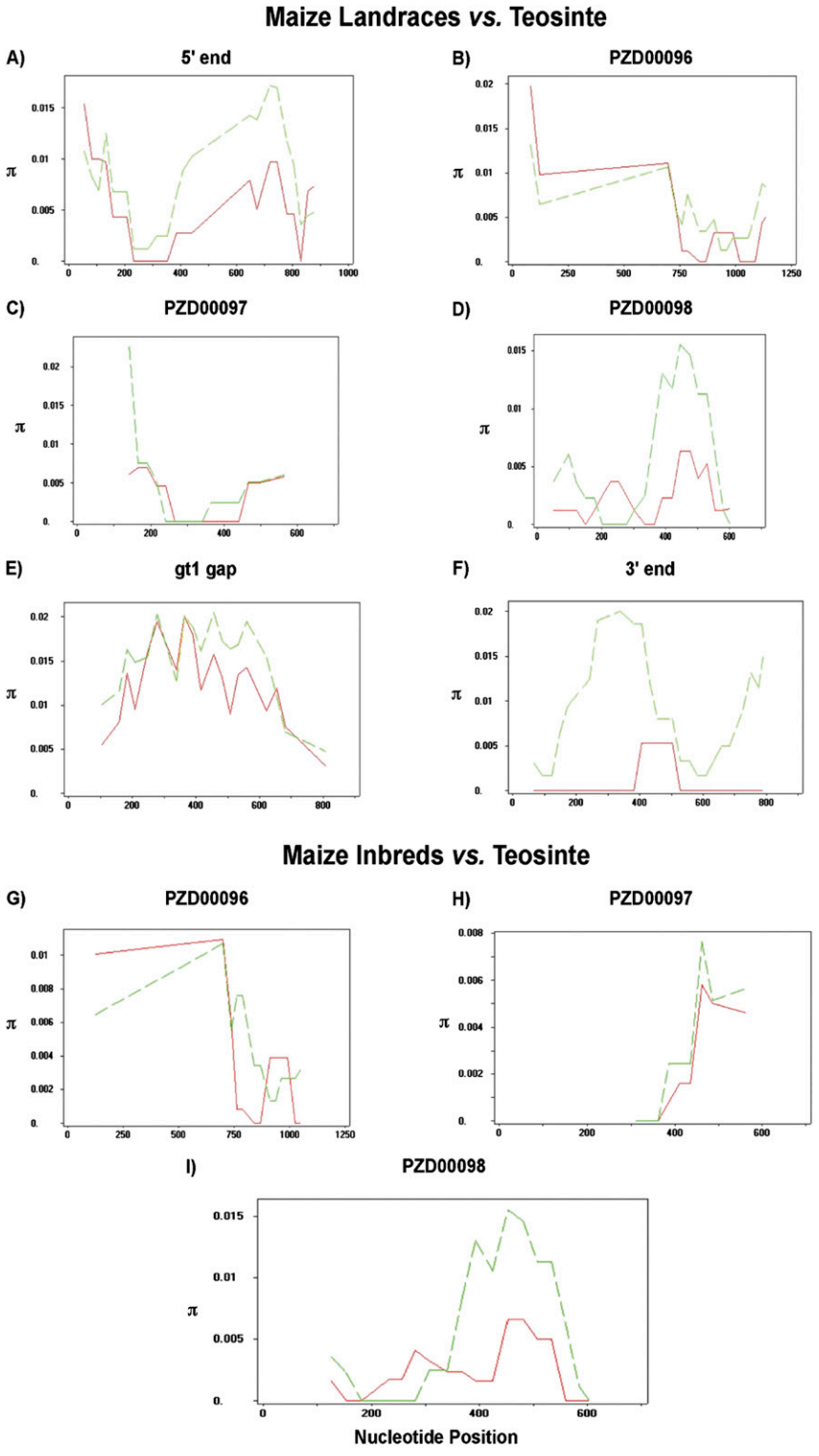

Fig. S2. Comparison of nucleotide diversity. The red line represents nucleotide diversity measured in the maize landraces, and the dashed green line represents nucleotide diversity in the teosinte individuals: 5′ end (*A*), PZD00096 (*B*), PZD00097 (*C*), PZD00098 (*D*), gt1 gap (*E*), and 3′ end (*F*). The red line represents nucleotide diversity measured in the maize inbreds, and the dashed green line represents nucleotide diversity in the teosinte individuals: PZD00096 (*G*), PZD00097 (*H*), and PZD00098 (*I*).

Table S1. Plant materials

Taxon	Accession no.	Origin
Z. mays ssp. *mays* (inbred)	B73	Iowa
	B97	Iowa
	CML52	Mexico
	CML69	Mexico
	CML103	Mexico
	CML228	Mexico
	CML247	Mexico
	CML277	Mexico
	CML322	Mexico
	CML333	Mexico
	HP301	Indiana
	IL14H	Illinois
	Ky21	Kentucky
	M37W	South Africa
	M162W	South Africa
	Mo17	Missouri
	Mo18W	Missouri
	NC350	North Carolina
	NC358	North Carolina
	Oh7B	Ohio
	Oh43	Ohio
	P39	Indiana
	Tx303	Texas
	Tzi8	Nigeria
	W22	Wisconsin
Z. mays ssp. *mays* (landrace)	PI213793	Northern United States
	OAX68	Southern Mexico
	URGII	Uruguay
	MEX48	Central Mexico
	SIN2	Western Mexico
	PUE32	Central Mexico
	VEN453	Venezuela
	CHI349	Chile
	GUA131	Guatemala
	CHH160	Northern Mexico
	MAG450	Columbia
	YUC7	Southern Mexico
	APC13	Peru
	SAN329	Columbia
	GUA14	Guatemala
	OXA70	Southern Mexico
Z. mays ssp. *parviglumis*	JSG y LOS 130	Tzitzio, Michoacan, Mexico
	JSG y LOS 119	Teloloapan, Guerrero, Mexico.
	JSG y MAS 401	La Lima, Jalisco, Mexico
	CIMMYT-8783	Teloloapan, Mexico
	JSG 197	Oaxaca, Mexico
	JSG y LOS 109	Palo Blanco, Mexico
	JSG 378	Acapulco, Mexico
	JSG 374	Tepoztlan, Morelos, Mexico
	JSG y LOS 161	Tejupilco, Mexico
	CIMMYT-11355	Teloloapan, Mexico
	PI566686	Huitzuco, Guerrero, Mexico
	JSG y MAS 264	Nayarit, Mexico
	Benz 967	El Rodeo, Mexico
	Kato site 4	Palo Blanco, Mexico
	Beadle and Kato site 4	Palo Blanco, Mexico
	Wilkes site 6	Teloloapan, Mexico
Z. diploperennis	Iltis et al., 1250	Las Joyas, Jalisco, Mexico
	R. Guzman 1120	Jalisco, Mexico
T. dactyloides	David Timothy Collection 68-14-1	Trujillo, Venezuela
	David Timothy Collection 68-23-1	Tachira, Venezuela

See www.panzea.org for more detailed information on these materials.

Table S2. Summary statistics

Amplicon	Maize inbreds				Maize landraces				Teosinte					
	L*	$S^{\dagger}$	$\pi^{\ddagger}$	$\theta^{\S}$	L*	$S^{\dagger}$	$\pi^{\ddagger}$	$\theta^{\S}$	L*	$S^{\dagger}$	$\pi^{\ddagger}$	$\theta^{\S}$	$\pi_{MI}/\pi_T{}^{\P}$	$\pi_{ML}/\pi_T{}^{\|}$
5′ end	NA	NA	NA	NA	677	12	0.00579	0.00571	669	27	0.00775	0.01216	NA	0.74709
PZD00096	394	4	0.00376	0.00272	500	8	0.00575	0.00482	497	18	0.00696	0.00859	0.54023	0.82615
PZD00097	470	7	0.00251	0.00399	470	8	0.00362	0.00513	466	7	0.00648	0.00528	0.38735	0.55864
PZD00098	544	6	0.00228	0.00299	597	7	0.00211	0.00353	593	13	0.00512	0.00661	0.44531	0.41211
gt1_gap	NA	NA	NA	NA	574	23	0.01095	0.01260	541	31	0.01402	0.01802	NA	0.78103
3′ end**	NA	NA	NA	NA	677	1	0.00079	0.00045	652	23	0.00853	0.01168	NA	0.09261

*Number of total sites, excluding gaps. NA (not available).
†Number of segregating sites (*S*), measured for total sites.
‡Nucleotide diversity.
§Nucleotide polymorphism.
¶Ratio of nucleotide diversity in the maize inbreds and the teosinte individuals.
‖Ratio of nucleotide diversity in the maize landraces and the teosinte individuals.
**Only the segment of the alignment with outgroup sequence available was used for calculations.

Table S3. Tajima's *D*-statistic

Amplicon	MI	MH	TI
5′ end	NA	0.05925	−1.96273* (−1.49204)
PZD00096	1.04868	0.69395	−0.75485
PZD00097	−0.67050	−1.06308	1.44905 (0.83622)
PZD00098	−0.72390	−1.4218	−0.87061
gt1_gap	NA	−0.55269	−0.95487
3′ end	NA	1.50272	−1.19650

Values in parentheses are Tajima's *D*-statistics calculated using η as opposed to *S*. For values, where there is no value in parentheses, Tajima's *D*-statistics were equivalent when calculated with η and segregating sites. MI, Maize Inbred Lines; MH, Maize Land Races; TI, Teosinte Inbreds.
*Significant value ($P < 0.5$) is indicated.

Table S4. Results of the HKA and coalescence simulation tests

Amplicon	*P* values from the HKA test*			*P* values from coalescence simulation test (maize landraces vs. teosinte)
	Maize inbreds	Maize landraces	Teosinte	
5′ end	NA	0.6312 (0.9604)	0.8294 (0.9999)	0.5293
PZD00096	0.2765	0.6686 (0.9747)	0.7984 (0.9999)	0.6565
PZD00097	0.8121	0.7979 (0.9897)	0.3050 (0.9600)	0.8789
PZD00098[b†]	0.9836	0.9705	0.9856	0.5690
gt1 gap	NA	0.9949 (0.9829)	0.9621 (0.9976)	0.7546
3′ end	NA	<<0.001 (<<0.001)	0.9528 (0.9990)	0.0219

*Values in parentheses are those calculated using the 15 control genes that previous analysis had indicated were neutral (1).
†*Z. diploperennis* as opposed to *T. dactyloides* was used as the outgroup.

1. Zhao QA, et al. (2008) The role of regulatory genes during maize domestication: Evidence from nucleotide polymorphism and gene expression. *Genetics* 178:2133–2143.

Other Supporting Information Files

Dataset S1 (TXT)