The Molecular Evolution of *terminal ear1*, a Regulatory Gene in the Genus Zea

Shawn E. White and John F. Doebley

Department of Plant Biology, University of Minnesota, Saint Paul, Minnesota 55108

Manuscript received January 2, 1999
Accepted for publication July 28, 1999

ABSTRACT

Nucleotide diversity in the *terminal ear1* (*te1*) gene, a regulatory locus hypothesized to be involved in the morphological evolution of maize (*Zea mays* ssp. *mays*), was investigated for evidence of past selection. Nucleotide polymorphism in a 1.4-kb region of *te1* was analyzed for a sample of 26 sequences isolated from 12 maize lines, five populations of the maize progenitor, *Z. mays* ssp. *parviglumis*, six other Zea populations, and two Tripsacum species. Although nucleotide diversity in *te1* in maize is reduced relative to ssp. *parviglumis*, phylogenetic and statistical analyses of the pattern of polymorphism among these sequences provided no evidence of past selection, indicating that the region of the gene studied was probably not involved in maize evolution. The level of reduction in genetic diversity in *te1* in maize relative to its progenitor is comparable to that found in previous reports for isozymes and other neutrally evolving maize genes and is consistent with a genome-wide reduction of genetic diversity resulting from a domestication bottleneck. An estimate of the age (1.2–1.4 million yr) of the maize gene pool based on *te1* is roughly consistent with previous estimates based on other neutral genes, but may be biased by the apparently slow synonymous substitution rate at *te1*.

THE crop plant maize (*Zea mays* ssp. *mays*) is thought to have been domesticated ~5000–10,000 yr ago from the wild annual plant, teosinte (*Z. mays* ssp. *parviglumis*; Iltis 1983). Several traits distinguish the two organisms, most of which involve the morphology of lateral branches and the inflorescences borne on them. Teosinte plants have long lateral branches tipped with male inflorescences (tassels), and small ears bearing kernels encased in hard stone-like fruitcases. Maize plants have very short lateral branches tipped with large ears bearing exposed kernels. Quantitative trait locus (QTL) mapping experiments have determined that these morphological differences between maize and teosinte are primarily due to five chromosomal regions (Doebley *et al.* 1990; Doebley and Stec 1993), each with major effects on one or more of these traits.

One of the QTL identified in these mapping experiments, found on the long arm of chromosome 3 (3L), has an effect on several traits distinguishing maize and teosinte. Teosinte plants into which the maize allele of the QTL on 3L has been introgressed via backcross breeding display a greater number of seeds per ear, a larger number of shorter lateral branch internodes, and feminized lateral inflorescences (Doebley *et al.* 1995). Two maize genes mapping to 3L [*tassel-replaces-upper-ear1* (*tru1*) and *terminal ear1* (*te1*)] have been identified as candidates for this QTL (Doebley *et al.* 1995). The *tru1* mutant allele of maize conditions a phenotype characterized by elongated upper lateral branches tipped by tassels instead of ears, while mutant alleles of *te1* cause an increase in the frequency of leaf primordia initiation and the feminization of the terminal inflorescence on the main stalk (Veit *et al.* 1993, 1998). *te1* is a regulatory gene that encodes a protein with conserved RNA-binding domains; it may function through RNA-binding activity (Veit *et al.* 1998). Both are attractive candidates for the QTL on 3L since *tru1* affects the fate of axillary meristems and *te1* affects the pattern of internode initiation, as does the QTL.

To begin to evaluate which of these two genes, if either, was involved in the origin of maize, the pattern of DNA sequence polymorphism in *te1* in maize and teosinte was analyzed for evidence of past selection. Although *te1* nucleotide diversity is reduced in maize relative to its progenitor, no evidence for selection during maize evolution was detected for the region analyzed. The neutral pattern of evolution at *te1* was exploited to investigate the age of the maize gene pool, the time of the divergence of the Zea and Tripsacum lineages, and the strength of the proposed domestication bottleneck.

MATERIALS AND METHODS

PCR amplification, cloning, and sequencing: An ~1.4-kb segment of the *te1* gene was PCR amplified from 25 Zea and Tripsacum individuals (Table 1). A single *te1* sequence from 22 Zea and 2 Tripsacum plants, and both alleles from 1 *Z. mays* ssp. *mexicana* (IL769) plant, were amplified and sequenced to yield a total of 26 *te1* sequences. The primers (TACAG CCGCTTCCGCAACAG and TGACGGTGGTCCTCGTATCC) used in the amplifications were designed from the cDNA se-

Corresponding author: John Doebley, Genetics Department, University of Wisconsin, 445 Henry Mall, Madison, WI 53706.
E-mail: jdoebley@facstaff.wisc.edu

TABLE 1

Zea and Tripsacum lines used in this study

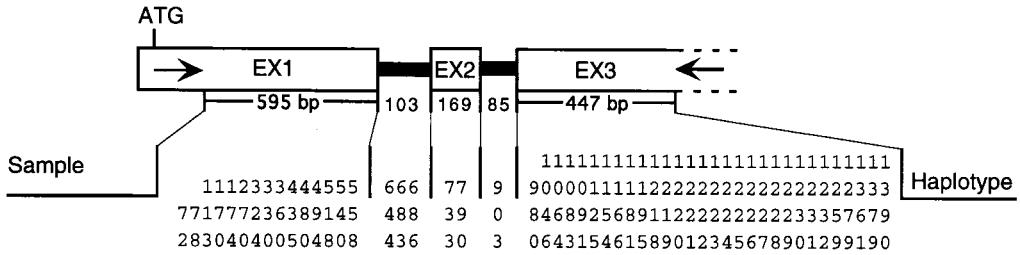
Species	Race/population	Accession no.	Location	Source[a]	Sample
ssp. *mays*	U.S. inbred	A619	United States	MAES	maize1
	U.S. inbred	A682	United States	MAES	maize2
	U.S. inbred	MA401	United States	WT	maize3
	U.S. inbred	W22	United States	JK	maize4
	Zapalote Grande	Chis224	Chiapas, Mexico	CIMMYT	maize5
	Chapalote	Sin2	Sinaloa, Mexico	CIMMYT	maize6
	Nal-Tel	Yuc7	Yucatan, Mexico	CIMMYT	maize7
	Reventador	Nay15	Nayarit, Mexico	CIMMYT	maize8
	Canilla Venezolano	Ven604	Venezuela	MG	maize9
	Pira	Ven485	Venezuela	MG	maize10
	Enano Gigante	Ecu969	Ecuador	MG	maize11
	Coroico	Bov396	Bolivia	MG	maize12
ssp. *parviglumis*	Teloloapan-Arcelia	81	Guererro, Mexico	HI & TC	parv1
	El Salado	—	Guererro, Mexico	GB	parv2
	El Rodeo	28888b	Jalisco, Mexico	HI	parv3
	Tzitzio	308	Michoacan, Mexico	HI & TC	parv4
	Teloloapan	C-11-78	Guererro, Mexico	CIMMYT	parv5
ssp. *mexicana*	Chalco-Amecameca	178	Mexico, Mexico	HI & TC	mex1
	Los Reyes	IL769	Mexico, Mexico	HI & AL	mex2a
	Los Reyes	IL769	Mexico, Mexico	HI & AL	mex2b
ssp. *huehuetenangensis*	San Antonio Huista	G-120	Guatemala	HI	hue1
	San Antonio Huista	G-119	Guatemala	HI	hue2
Z. diploperennis	Manantlan	2549	Jalisco, Mexico	HI	—
Z. luxurians	Agua Blanca	G-38	Jutiapa, Guatemala	HI	—
T. dactyloides	—	—	—	—	—
T. floridanum	—	79-35	Florida	LR	—

[a] AL, Alex Lasseigne; CIMMYT, Centro Internacional para Mejoramiento de Maiz y Trigo; GB, George Beadle; HI, Hugh Iltis; JK, Jerry Kermicle; LR, Lawrence Randolph; MAES, Minnesota Agricultural Experiment Station; MG, Major Goodman; TC, Ted Cochrane; WT, William Tracy.

quence of the *te1* gene (Veit *et al.* 1998), and correspond to the 5′ half of the cDNA. PCR amplifications were performed using PCR Supermix (BRL), 10% DMSO, and 10 pg of each primer under the following conditions: 30 cycles of 95°, 1′; 57°, 1′; 75°, 1′ followed by a 10′ final extension step at 75°. The products of these amplifications were cloned into the TA vector (Invitrogen, San Diego) and sequenced, either individually or in pools (see below). Sequencing was performed by an automated sequencer at the University of Minnesota Micro-Chemical Facility (Minneapolis) or the University of Minnesota Advanced Genetic Analysis Center (St. Paul).

Isolation of single *te1* alleles: Many of the individuals used in this study were members of noninbred strains, and therefore likely to be heterozygous at *te1*. PCR products amplified from heterozygous individuals would represent both alleles, and if pooled prior to sequencing (to reduce the contribution of errors introduced by *Taq* polymerase), would yield ambiguous sequencing results. Therefore, DNA restriction fragments with single *te1* alleles were isolated prior to PCR amplification. This was done by identifying heterozygous individuals via Southern blot analysis of DNA digested with restriction enzymes that do not cut *te1*, excising DNA fractions containing single alleles from low-melting-point agarose gels, isolating the DNA from gel slices using Geneclean (Bio101), and using this DNA as the substrate for PCR. Ten positive clones of PCR products of each *te1* allele obtained by this method were pooled before sequencing. Additional sequences from some samples were obtained using total cellular DNA from individual plants as the substrate for PCR and then cloning and sequencing a single PCR product. This method was employed for Tripsacum and *Z. mays* ssp. *huehuetenangensis*, which were not used in assays of nucleotide polymorphism, and for maize inbred lines. In the case of the inbred lines, sequences obtained from single clones were inspected for "singletons," polymorphisms found only in that sequence relative to all other sequences. These differences can represent either errors introduced into the sequence by *Taq* polymerase or true sequence variation. Reamplification and partial resequencing of these inbreds indicated that all singletons were due to *Taq* error and so the corrected sequences were used in the analyses.

Data analysis: A phylogenetic tree based on the *te1* sequences was constructed using the neighbor-joining method (PHYLIP version 3.5c, Felsenstein 1993) and Kimura two-parameter distances (Kimura 1980). The level of support for branch points in the tree was assessed by determining the consistency of the branch points among 200 bootstrap resamplings of the original data. Two measures of nucleotide diversity were calculated using the SITES program (Hey and Wakeley 1997): π, the expected heterozygosity per nucleotide site (Nei 1987), and $\hat{\theta}$, an estimate of $4N_e\mu$, where N_e is the effective population size and μ is the mutation rate per nucleotide (Watterson 1975). Evidence for past selection was investigated using the Hudson, Kreitman, Aguadé (HKA) test (Hudson *et al.* 1987), the Tajima test (Tajima 1989), and the McDonald-Kreitman (M-K) test (McDonald and Kreitman 1991). The HKA test was performed using the method of Hilton *et al.* (1994) with the *alcohol dehydrogenase1* (*adh1*) and *adh2* sequence data sets (Gaut and Clegg 1993a; Goloubi-




```
Consensus      GGCGTCCCCCGCAGC  GGC  GC  C  GGGGATGCGGGCGGCGGCG------GTCGC
maize1         ......T........  ...  ..  .  ..................A.........A.  1
maize2         ...............  ...  ..  .  T........C.....................  2
maize3         ......T........  ...  ..  .  ..................A.........A.  1
maize4         ......T........  ...  ..  .  ..................A.........A.  1
maize5         ......T........  ...  ..  .  ..................A.........A.  1
maize6         ....A..........  ...  A.  .  ...............................  3
maize7         ......T........  ...  ..  .  ..................A.........A.  1
maize8         ...............  ...  ..  .  T........C.....................  2
maize9         ......T........  ...  ..  .  ..................A.........A.  1
maize10        ......T........  ...  ..  .  ..................A.........A.  1
maize11        ...............  ...  ..  .  T........C.....................  2
maize12        ...............  ...  ..  .  T........C.....................  2
parv1          .....T..T.C....  ...  ..  .  ..................A.........A.  4
parv2          .T.............  ...  ..  .  ............................A....  5
parv3          .T.............  ...  ..  .  ..................GCGGCG....T  6
parv4          ...........G...  ...  ..  .  ...............................  7
parv5          .....A.........  ...  AA  .  ...............................A.  8
mex1           ...............  ...  ..  .  T........C.....................  2
mex2a          .......G.TC....  ...  ..  .  .....G.............GCG........  9
mex2b          A............CA  ...  AA  .  ........A.---------.........A.  10
hue1           ...A....T.C.G..  ..T  ..  .  ..AA...T...................G..  11
hue2           ........T.C....  ...  ..  .  ...A...T...................G..  12
diplo          ...............  .A.  ..  G  .A...G....---------...........  13
lux            ..A....G..C....  A..  ..  .  ....GGA......------.......C...  14

Synon/Nonsyn   SSNNSSSSSSSNNSS       SS     SSSNSNSSSN        S      SNSSN
```

Nonsynonymous mutations:

```
113 THR to ASN (polar to polar)
170 GLY to ASP (polar to acidic)
494 ALA to GLY (nonpolar to polar)
518 ASN to SER (polar to polar)
1083 ALA to THR (nonpolar to polar)
1125 SER to ALA (polar to nonpolar)
1195 GLY to ALA (polar to nonpolar)
1279 VAL tp ALA (nonpolar to nonpolar)
1390 ALA to VAL (nonpolar to nonpolar)
```

Figure 1.—A schematic diagram of the *te1* region analyzed in this study and the nucleotide polymorphisms found in that region among Zea sequences. Arrows indicate the position of PCR primers used, and open boxes represent exons 1 (EX1), 2 (EX2), and 3 (EX3). The sample designations are given in Table 1. Position numbers refer to the genomic region of *te1* between the primers used. Dots indicate sequence identity to the consensus sequence, and dashes represent indels. The synonymous (S) and nonsynonymous (N) changes and the nature of the nonsynonymous changes are indicated.

noff *et al.* 1993) included as controls. The Tajima test and the M-K test were performed using DnaSP (Rozas and Rozas 1997). Where required for these tests, we used *Z. diploperennis* as an outgroup. Recombination was assessed using the algorithm of Hudson and Kaplan (1985) in the SITES program. The numbers of synonymous and nonsynonymous substitutions per site were determined by the method of Nei and Gojobori (1986).

RESULTS

Nucleotide polymorphism at *te1*: To assay nucleotide polymorphism in *te1*, we determined the genomic sequence of a 1399-bp region of the *te1* gene, including two small introns for 24 Zea samples (Figure 1). Over a total sequence length of 1399 bp, 1211 bp of which are coding, there are 37 polymorphisms across all Zea species (Figure 1). Four of the polymorphic sites are found within intron sequences, giving 2.1% polymorphic nucleotide positions in introns, while 33 sites are polymorphic within exons, giving 2.7% polymorphic sites in the coding region. Of 309.67 synonymous sites, 24 (7.8%) were polymorphic, as were 9 of 901.33 nonsynonymous sites (1.0%). Thus, although the introns are more polymorphic than the nonsynonymous sites, their level of polymorphism is closer to that of the nonsynonymous sites than to the synonymous sites. Of the nine nonsynonymous substitutions, four result in nonconservative amino acid changes (Figure 1). There is one insertion/deletion (indel) of variable size that is due to the presence of different numbers of the short repeat "GCG." This indel is fixed in size among the maize *te1* sequences, but exhibits length variation among the other taxa of Zea.

Among the 24 Zea sequences, there are 14 haplotypes

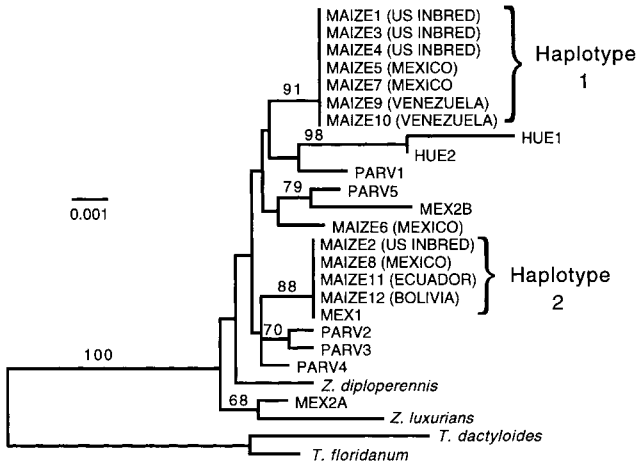


Figure 2.—A neighbor-joining tree of the *te1* sequences. Bootstrap values >50% are shown. Branch lengths are proportional to the probable number of substitutions per site using Kimura 2-parameter distances.

(Figure 1). With the exception of 1 sequence (maize6), all of the maize *te1* sequences fall into 1 of 2 haplotypes. There is no correlation between these 2 haplotypes and the geographical source of the maize lines sampled (Table 1). One of these 2 haplotypes was found only among maize individuals, while the other was found in maize and 1 teosinte (*Z. mays* ssp. *mexicana*) individual (mex1). In the remaining 11 Zea individuals, 11 haplotypes were found.

Haplotype diversity is influenced by the amount of recombination at a locus. We assessed the minimum number of recombination events required to explain the observed patterns of segregating sites within each subspecies included in the *te1* data set using the method of Hudson and Kaplan (1985). This test shows no evidence of recombination within either maize or ssp. *parviglumis* ($R_m = 0$). However, when these two taxa are considered together, there is evidence for at least one recombination event ($R_m = 1$).

Phylogenetic analysis of *te1* sequences: If *te1* were involved in the evolution of maize, selection for a specific sequence motif at this locus could be expected to result in a *te1* gene tree in which all maize sequences form a single clade. To test this prediction, phylogenetic analysis was performed on the *te1* sequence data. The resulting phylogenetic tree does not reveal a monophyletic maize clade (Figure 2). In fact, the tree shows little resolution for any of the Zea species or subspecies. In many cases a sequence from one subspecies is more closely related to sequences from another subspecies than it is to sequences from the same subspecies. A similar lack of resolution is found in phylogenetic trees made using other Zea nuclear genes (Gaut and Clegg 1993a; Goloubinoff *et al* 1993; Buckler and Holtsford 1996; Hanson *et al.* 1996; Hilton and Gaut 1998), and is probably indicative of lineage sorting and/or introgression among Zea taxa. The two *Z. mays* ssp. *huehuetenangensis* sequences included in the analysis, however, group together in the tree.

TABLE 2

Nucleotide diversity (π and $\hat{\theta}$) at Zea loci

		Maize		ssp. *parviglumis*	
Locus	Sites	π	$\hat{\theta}$	π	$\hat{\theta}$
te1	Total	2.08	1.66	3.66	3.79
	Silent	4.85	4.00	8.65	8.69
	Synonymous	7.80	6.43	13.9	13.97
adh1[a]	Total	13.60	14.91	19.5	17.91
	Silent	21.59	18.1	23.4	24.5
	Synonymous	28.0	21.8	51.0	51.4
adh2[b]	Total	24.95	25.76	32.79	32.79
	Silent	30.06	29.47	39.55	39.55
c1[c]	Total	2.67	3.50	12.17	12.21
	Silent	2.84	4.11	15.04	14.98
glb1[d]	Total	17.23	18.88	25.81	31.72
	Silent	17.95	18.98	27.25	23.15
	Synonymous	37.22	36.16	45.20	55.61
o2[e]	Silent	—	52.2	—	—
tb1[f]	5′ flanking	0.47	0.93	28.68	32.57
238[g]	Total	—	40.0	—	—
288[g]	Total	—	27.0	—	—
451[g]	Total	—	13.0	—	—
455[g]	Total	—	6.0	—	—

Values per basepair ×1000. Silent sites were defined as 5′ leader, intron, and third position sites.

[a] Eyre-Walker *et al.* (1998).

[b] Goloubinoff *et al.* (1993). Synonymous sites not calculated because there are few sites.

[c] Hanson *et al.* (1996). Synonymous sites not calculated because there are few sites.

[d] Hilton and Gaut (1998).

[e] Henry and Damerval (1997).

[h] Wang *et al.* (1999).

[g] Random loci; Shattuck-Eidens *et al.* (1990).

Comparison of *te1* nucleotide diversity with other Zea loci: Values for nucleotide diversity at *te1* in maize are among the lowest of the Zea loci for which π and $\hat{\theta}$ have been estimated (Table 2). The only genes with less diversity [*teosinte branched1* (*tb1*) and *colorless aleurone* (*c1*)] are known or suspected to have been under selection in maize. Nucleotide diversity at *te1* is also low in ssp. *parviglumis*. Maize π and $\hat{\theta}$ values for *te1* at total sites are ~57 and 44% of these values for ssp. *parviglumis*. At other loci, the percentage of variation found in maize compared to ssp. *parviglumis* ranges from 2.9% for *tb1*, to 60% for *globulin1* (*glb1*) and 83% for *adh1* when total sites are considered (Table 2; Hilton and Gaut 1998).

Tests for selection at *te1*: Low nucleotide diversity may indicate that selection has reduced polymorphism at a locus. To test whether this is the case for *te1* in maize, the HKA test was performed on the maize *te1* sequence data using *Z. diploperennis* as the outgroup. This test assays whether there is a lack of consistency in the ratio of polymorphism to divergence between unlinked loci. When applied to *te1* using *adh1* and *adh2*

sequences as controls, the HKA test did not reject the null hypothesis that *te1* is evolving neutrally in maize ($X^2 = 0.626$, $P = 0.73$). Therefore, although nucleotide diversity at *te1* is low compared to other maize loci, similarly low *te1* divergence values between species suggest that this low level of polymorphism is not due to positive selection within the maize lineage.

Another common test for departure from neutrality, the Tajima test, uses a comparison of π and θ to detect selection (Tajima 1989). Under neutral evolution, π and θ are expected to be equal, so the Tajima *D* statistic, $\pi - \hat{\theta}/[V(\pi - \hat{\theta})]^{-1/2}$, will be zero when a locus is evolving neutrally, and significantly different from zero when selection has affected the frequencies of the polymorphisms. There are only 3 maize haplotypes among 12 maize sequences compared to 12 haplotypes among 12 teosinte sequences. So at least superficially, it appears that selection may have affected the pattern of polymorphism in maize. However, the *D* value for the maize *te1* sequence data ($D = 0.97$) is not significantly different from zero ($P > 0.1$), and the null hypothesis that the pattern of polymorphism is due to neutral processes cannot be rejected.

The relative numbers of synonymous and nonsynonymous nucleotide substitutions within and between species can also be used to detect selection using the M-K test. This test assumes that under neutral processes, the ratio of replacement to synonymous fixed differences between species will be the same as the ratio of replacement to synonymous polymorphisms within species. When applied to the maize *te1* data using *Z. diploperennis* for interspecies comparison, the M-K test revealed no departure from neutrality (G value with William's correction = 0.588, $P = 0.44$).

Divergence dates: Since *te1* appears to be evolving neutrally, nucleotide divergences at *te1* should be appropriate for determining the time of historical events such as the age of the maize gene pool and the divergence time of the Zea-Tripsacum lineages. The two most divergent maize haplotypes (haplotypes 1 and 3) differ by five synonymous substitutions over 309.67 synonymous sites (Figure 1). Assuming that the substitution rate in maize is 5.9 to 6.5×10^{-9} substitutions per synonymous site per year for neutral genes (Gaut *et al.* 1996; see discussion), the maize gene pool is ~1.2–1.4 million years (myr) old. The three maize *te1* haplotypes therefore diverged well before maize was domesticated (~7500 yr ago; Iltis 1983), and were inherited by the maize lineage from an ancestral population. Similarly, using net divergence (Nei 1987) at synonymous sites between the two Tripsacum species and Zea (9.33/308 sites) and the above mutation rates, the Zea and Tripsacum lineages diverged ~2.3–2.6 mya.

DISCUSSION

Phylogeny: If *te1* were involved in the evolution of maize morphology, then the selected sequence difference(s) would act as synapomorphies that could join all maize sequences in a single clade apart from other Zea. Phylogenetic analysis of the Zea and Tripsacum *te1* sequences, however, reveals no evidence for a monophyletic maize clade. Rather, the maize sequences fall primarily into two groups, one of which includes a sequence from ssp. *mexicana* (mex1). This ssp. *mexicana* sequence is of the same haplotype as many of the maize sequences, possibly due to introgression between ssp. *mexicana* and maize. Introgression between these two subspecies has been reported previously (Doebley 1990b). Alternatively, both maize and ssp. *mexicana* may have retained an ancient haplotype that was present in an ancestral population. Similarly, maize shares three polymorphisms with the ssp. *parviglumis* sequences, and two with other ssp. *mexicana* sequences. While these shared polymorphisms could also reflect introgression between the three taxa, a more likely explanation is the retention of ancient polymorphisms by each of these *Z. mays* subspecies as has been previously suggested for *adh1* and *colored aleurone1* (*c1*; Gaut and Clegg 1993a; Hanson *et al.* 1996). Since the topology of the *te1* tree for Zea is similar in nature to that of neutral genes such as *adh1*, it offers no support for the hypothesis that *te1* has been involved in maize morphological evolution.

No evidence for selection on *te1*: Three statistical tests designed to determine whether a locus is evolving neutrally all failed to show evidence for past selection acting on the maize *te1* gene. One of the predictions made by these tests is that if *te1* were involved in maize evolution, the relative level of nucleotide diversity at *te1* in maize should be reduced as compared to neutral genes such as *adh1* and *adh2*. However, the HKA test, when applied to sequence data for *te1* and the appropriate controls, shows no evidence for a reduction in nucleotide diversity in *te1* relative to these other two genes. Similarly, the McDonald-Kreitman test, which compares the ratio of replacement to synonymous fixed differences with the ratio of replacement to synonymous polymorphisms, failed to show any evidence of selection at *te1*. The third test, the Tajima test, asks whether π and θ are equivalent as expected under neutral evolution. When applied to the maize *te1* data, the Tajima test did not reject the null hypothesis that *te1* is evolving in a neutral manner.

Despite the fact that all of the statistical tests for selection failed to show any evidence for selection, a role for the *te1* locus in maize evolution cannot be completely discounted. These tests are designed to be conservative and so they reject the null hypothesis only in very obvious cases of selection. Selection at *te1* may therefore be too weak to be detected by these tests. Alternatively, if the region of *te1* under selection in maize is upstream or downstream of the region analyzed, these tests would not detect selection if the recombination rate within the locus is high enough to allow closely linked regions to have different histories. Similar statistical analysis of the effects of selection on the *tb1* locus of maize has

shown that recombination can be sufficiently high that neighboring sequences can have very different histories (Wang *et al.* 1999). Specifically, comparison of the *tb1* 5′ flanking region with the coding region of the gene indicated that selection has impacted the pattern of polymorphism in one region but not the other. Thus, if selection acted upon regions of the *te1* gene other than the one analyzed, it could have gone undetected. Therefore, the results presented here suggesting that *te1* was not involved in maize evolution should be considered preliminary.

Nucleotide diversity in *te1* is low: The level of nucleotide diversity in *te1* is among the lowest of any maize locus and *te1* has 10-fold less diversity than some other maize loci (Table 2). The only genes with less diversity (*tb1* and *c1*) are known or suspected to have been under selection in maize. Moreover, diversity for *te1* is lower than the reported values for most other genes whether one considers total, silent, or synonymous sites (Table 2). As suggested by the nonsignificant result of the HKA test, the low diversity of *te1* in maize is likely a characteristic of this gene and not the result of a selective sweep. Consistent with this interpretation, a low level of diversity for *te1* appears to be maintained not only in maize but also in ssp. *parviglumis* for which *te1* has the lowest value for $\hat{\theta}$ observed in this subspecies. That selection in maize is unlikely to explain the low diversity for *te1* is made particularly clear if one compares *te1* to *tb1*. At *tb1* where there is evidence for selection in maize, diversity is low in maize but high in ssp. *parviglumis*. If one applies the HKA test to *tb1* and *te1*, the result is highly significant. This result is consistent with our interpretation that *tb1* has experienced a selective sweep while *te1* has not (data not shown).

There are at least three explanations for the low level of nucleotide diversity in *te1*. First, the substitution rate at *te1* may be lower than it is for other maize genes. Synonymous substitution rates have been found to vary among plant nuclear genes (Wolfe *et al.* 1989) and may indicate that different genes experience different mutation rates. Second, background selection on *te1*, itself due to a high degree of functional constraint or selection on neighboring genes coupled with a hitchhiking effect, could reduce variation at *te1* (Maynard Smith and Haigh 1974; Charlesworth *et al.* 1993). If such selection were acting in the progenitor of extant Zea species, there would have been less variation to partition into its descendants. Third, codon usage bias can bring synonymous sites under selective constraint. Codon usage bias appears to vary among mammalian genes and causes a reduced synonymous substitution rate in some genes relative to others (Li 1997).

Previously, isozyme data were used to ask whether bottlenecks during domestication caused a reduction in polymorphism in crops relative to their progenitors (Doebley 1989). The isozyme data indicate that an average crop possesses ~75% of the genetic diversity (expected heterozygosity) and 80% as many polymorphic loci as its progenitor. Isozyme data, based on 21 loci, indicated that maize possesses only 70% as much heterozygosity and only 73% as many polymorphic loci as its progenitor. More recent studies of nucleotide diversity at neutrally evolving loci are consistent with these results, demonstrating that maize possesses ~83% of the diversity found in its progenitor at *adh1* and ~60% at *glb1* (Hilton and Gaut 1998).

The percentage of diversity that maize has retained in *te1* relative to ssp. *parviglumis* is consistent with these previous reports, although somewhat smaller. The values of π and $\hat{\theta}$ for *te1* in maize are 57 and 44% of these values for ssp. *parviglumis*. In addition, there are only three maize haplotypes among 12 maize individuals compared to five haplotypes among 5 ssp. *parviglumis* individuals, suggesting that haplotypic diversity is reduced more severely than nucleotide diversity. Simulations of the bottleneck process have shown that the reduction in maize genetic diversity can be explained by a short bottleneck involving only a few individuals (Eyre-Walker *et al.* 1998). The low number of *te1* haplotypes in maize may therefore be due to a narrow bottleneck from which only a few haplotypes emerged.

Time of historical events: Since *te1* appears to be evolving in a neutral manner, *te1* sequences should be useful in estimating the time of historical events. To do this requires that the rate of sequence evolution at synonymous sites be known. Gaut and colleagues (Gaut *et al.* 1996; Eyre-Walker *et al.* 1998; Hilton and Gaut 1998) have used a rate of 6.5×10^{-9} substitutions per synonymous site per year based on sequence divergence at synonymous sites in *adh1* and *adh2* between maize and rice/barley and an estimated divergence date of 50 myr for these grasses. To examine the generality of this rate, we calculated the synonymous rate between maize and rice for five genes (*ant*, *c1*, *c2*, *cdc*, and *ohp*; see Gaut and Doebley 1997) and obtained a rate of 5.9×10^{-9} when these genes are averaged with *adh1* and *adh2*. The rates among these genes vary from 4.7×10^{-9} for *c2* to 7.0×10^{-9} for *adh1*. One potential source of error in these estimates is the presumed maize-rice divergence date of 50 myr. The fossil record does not provide an unambiguous estimate of the time of maize-rice divergence and various authors have placed the date in a range from 40–70 myr (Wolfe *et al.* 1989; Gaut *et al.* 1996). Another potential source of error is rate variation among grass lineages. Gaut and Clegg (1993b) estimated that the rate of *adh1* sequence evolution is 1.7 times higher in the maize lineage than in the pennisetum lineage. If this represents a recent acceleration of the rate in the maize lineage, applying a general grass rate to maize would bias the estimates of the time of historical events to be older than they actually are.

Gaut and Clegg (1993a) and Goloubinoff *et al.* (1993) first established that the age of the maize gene pool for neutral genes greatly exceeds the age of maize

as a species by calculating the divergence time for maize alleles at *adh1* and *adh2*, respectively. Gaut and Clegg (1993a) estimated that the most distinct *adh1* alleles diverged 1.9 mya. More recently, SanMiguel *et al.* (1998) estimated a divergence time of 3.6 myr for *adh1*, a value nearly twice the original estimate. (The difference between these estimates partly reflects application of the synonymous rate to intron sites in one of these studies.) Hanson *et al.* (1996) estimated the age of the maize gene pool at 140,000 yr on the basis of the *c1* gene and a substitution rate of 5×10^{-9}; however, this estimate is likely biased downward since *c1* was probably under selection that eliminated some haplotypes during maize domestication. We estimate that the divergence time of the most distinct maize *te1* alleles is ~1.2–1.4 myr. This estimate agrees roughly with prior estimates in that it indicates that the age of the maize gene pool greatly exceeds the age of maize as a species (5,000–10,000 yr); however, it is only one-half to one-third as large as some estimates. A potential explanation for this discrepancy is a slower substitution rate for *te1* relative to the *adh* genes as discussed above. Alternatively, if during the domestication bottleneck, maize population sizes were very small (Eyre-Walker *et al.* 1998), then one would expect considerable variation in the extent to which different neutral genes retained the diversity found in the progenitor of maize.

Tripsacum has long been recognized as the sister genus to Zea (Mangelsdorf and Reeves 1939). Hilton and Gaut (1998) have estimated the time of divergence of the Zea and Tripsacum lineages at 4.5 to 4.8 mya. Our estimate of 2.3–2.6 myr based on *te1* is considerably more recent and may be biased downward if *te1* is evolving at a slower rate than other maize genes as discussed above. Although these dates do not provide an estimate of the age of Zea since the taxonomic status of these lineages 4.7 mya is unknown, they can be viewed as an upper bound, presuming that Zea is not paraphyletic. A minimal estimate or lower bound for the age of Zea is provided by the divergence time of extant Zea species, presuming Zea is not polyphyletic. *Zea mays* and *Z. luxurians*, two of the most distinct members of Zea that potentially represent the earliest divergence among extant members of the genus (Doebley 1990a), are estimated to have diverged ~700,000 yr ago (Hilton and Gaut 1998), indicating that the genus should be at least this age. Among angiosperms, old genera can be traced back in the fossil record for 50–100 myr or more [P. Crane, personal communication (Oct. 27, 1998)]. For example, the extant genera Ceratophyllum is known from the Paleocene (*ca.* 60 mya), while the extant genera Alnus and Betula are known unequivocally from the Eocene (*ca.* 45 mya; Crane 1989; Herendeen *et al.* 1990). In this context, Zea, with its age currently estimated as between 0.7 and 4.7 myr, is a young genus.

Finally, we have presented three estimates of population or historical parameters using *te1* that are lower than those found with other neutral genes: smaller estimates of θ, a younger age for the maize gene pool, and a younger divergence time for the Zea-Tripsacum lineages. As discussed above, a variety of explanations can be offered for these discrepancies; however, one factor, a slower synonymous substitution rate for *te1*, can substantially explain all three discrepancies. If the synonymous substitution rate for *te1* is half of the average for the seven nuclear genes we used to calculate a general rate, then the estimates for all three parameters would be much closer to estimates from the neutral genes *adh1*, *adh2*, and *glb1*. For this reason, we suspect that *te1* is evolving at a reduced rate. Better estimates of the time of historical events will require a more complete fossil record, an investigation of rate variation among different lineages, an understanding of rate variation among genes, and reconsideration of applying synonymous rates to intron sites. Until these matters are resolved, it should be recognized that estimates for the time of historical events surrounding maize evolution, while interesting and fun, could easily be off by a factor of two or more.

From QTL to gene: The motivation for this study was to test whether *te1* corresponds to a QTL involved in maize evolution (Doebley *et al.* 1995). The logic was simple: if *te1* is this QTL, then it should show the signature of a past selective sweep in maize. The signature of selection was not observed. Unfortunately, this negative result is of limited value since selection on the *te1* promoter may not have impacted nucleotide polymorphism in the coding region that we analyzed (see Wang *et al.* 1999). Nevertheless, we feel that a population genetic approach to testing gene:QTL correspondence remains of value, especially if the region analyzed includes both coding and regulatory sequences. The value of a population genetic approach is in part that alternative approaches to testing gene:QTL correspondence such as gene transformation are more costly, more time consuming, and fraught with their own sets of difficulties.

The authors thank Bruce Veit for sharing unpublished maize *te1* cDNA sequence, Rong-Lin Wang for sharing unpublished *tb1* data and for helpful discussion, Peter Crane, Brandon Gaut, and Jonathan Wendel for comments on the manuscript, and Brandon Gaut for providing the *glb1* data. This work was supported by a National Science Foundation / Sloan postdoctoral fellowship to S.E.W. and a grant from the National Science Foundation to J.F.D.

LITERATURE CITED

Buckler, E. S., and T. P. Holtsford, 1996 *Zea* systematics: ribosomal ITS evidence. Mol. Biol. Evol. **13:** 612–622.

Charlesworth, B., M. Morgan and D. Charlesworth, 1993 The effects of deleterious mutations on neutral molecular variation. Genetics **134:** 1289–1303.

Crane, P. R., 1989 Early fossil history and evolution of the Betulaceae, pp. 87–116 in *Evolution, Systematics, and Fossil History of the Hamamelidae*, Vol. 2, '*Higher*' *Hamamelidae*, edited by P. R. Crane and S. Blackmore. Clarendon Press, Oxford.

Doebley, J., 1989 Isozymic evidence and the evolution of crop plants, pp. 165–191 in *Isozymes in Plant Biology*, edited by D. E. Soltis and P. S. Soltis. Dioscorides Press, Portland, OR.

Doebley, J., 1990a Molecular evidence and the evolution of maize. Econ. Bot. **44** (Suppl. 3): 6–27.

Doebley, J., 1990b Molecular evidence for gene flow among *Zea* species. Bioscience **40:** 443–448.

Doebley, J., and A. Stec, 1993 Inheritance of morphological differences between maize and teosinte: comparison of results of two F_2 populations. Genetics **134:** 559–570.

Doebley, J., A. Stec, J. Wendel and M. Edwards, 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.

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

Eyre-Walker, A., R. L. Gaut, H. Hilton, D. L. Feldman and B. S. Gaut, 1998 Investigation of the bottleneck leading to the domestication of maize. Proc. Natl. Acad. Sci. USA **95:** 4441–4446.

Felsenstein, J., 1993 PHYLIP (Phylogeny Inference Package) version 3.5c. Distributed by the author. Department of Genetics, University of Washington, Seattle.

Gaut, B. S., and M. T. Clegg, 1993a Molecular evolution of the *Adh1* locus in the genus *Zea*. Proc. Natl. Acad. Sci. USA **90:** 5095–5099.

Gaut, B. S., and M. T. Clegg, 1993b Nucleotide polymorphism in the *Adh1* locus of pearl millet (*Pennisetum glaucum*) (Poaceae). Genetics **135:** 1091–1097.

Gaut, B. S., and J. Doebley, 1997 DNA sequence evidence for the segmental allotetraploid origin of maize. Proc. Natl. Acad. Sci. USA **94:** 6809–6814.

Gaut, B. S., B. R. Morton, B. M. McCaig and M. T. Clegg, 1996 Substitution rate comparisons between grasses and palms: synonymous rate differences at the nuclear gene *Adh1* parallel rate differences at the plastid gene *rbcL*. Proc. Natl. Acad. Sci. USA **93:** 10274–10279.

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

Hanson, M. A., B. S. Gaut, A. O. Stec, S. I. Fuerstenberg, M. M. Goodman *et al.*, 1996 Evolution of anthocyanin biosynthesis in maize kernels: the role of regulatory and enzymatic loci. Genetics **143:** 1395–1407.

Henry, A.-M., and C. Damerval, 1997 High rates of polymorphism and recombination at the *opaque-2* locus in cultivated maize. Mol. Gen. Genet. **256:** 147–157.

Herendeen, P. S., D. H. Les and D. L. Dilcher, 1990 Fossil *Ceratophyllum* (Ceratophyllaceae) from the tertiary of North America. Amer. J. Bot. **77:** 7–16.

Hey, J., and J. Wakeley, 1997 A coalescent estimator of the population recombination rate. Genetics **145:** 833–846.

Hilton, H., and B. Gaut, 1998 Speciation and domestication in maize and its wild relatives: evidence from the *globulin-1* gene. Genetics **150:** 863–872.

Hilton, H., R. M. Kliman and J. Hey, 1994 Using hitchhiking genes to study adaptation and divergence during speciation within the *Drosophila melanogaster* species complex. Evolution **48:** 1900–1913.

Hudson, R. R., and N. L. Kaplan, 1985 Statistical properties of the number of recombination events in the history of a sample of DNA sequences. Genetics **111:** 147–164.

Hudson, R. R., M. Kreitman and M. Aguadé, 1987 A test for neutral molecular evolution based on nucleotide data. Genetics **116:** 153–159.

Iltis, H. H., 1983 From teosinte to maize: the catastrophic sexual transmutation. Science **222:** 886–894.

Kimura, M., 1980 A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J. Mol. Evol. **16:** 111–120.

Li, W.-H., 1997 *Molecular Evolution*. Sinauer Associates, Sunderland, MA.

Mangelsdorf, P. C., and R. G. Reeves, 1939 *The Origin of Indian Corn and Its Relatives.* Texas Agric. Exp. Sta. Bull. No. 574, College Station, TX.

Maynard Smith, J., and J. Haigh, 1974 The hitchhiking effect of a favorable gene. Genet. Res. **23:** 23–35.

McDonald, J., and M. Kreitman, 1991 Adaptive protein evolution at the *Adh* locus in *Drosophila*. Nature **351:** 652–654.

Nei, M., 1987 *Molecular Evolutionary Genetics.* Columbia University Press, New York.

Nei, M., and T. Gojobori, 1986 Simple methods for estimating the number of synonymous and non-synonymous nucleotide substitutions. Mol. Biol. Evol. **3:** 418–426.

Rozas, J., and R. Rozas, 1997 DnaSP version 2.0: a novel software package for extensive molecular population genetics analysis. Comput. Appli. Biosci. **13:** 307–311.

SanMiguel, P., B. S. Gaut, A. Tikhonov, Y. Nakajima and J. L. Bennetzen, 1998 The paleontology of intergene retrotransposons of maize. Nat. Genet. **20:** 43–45.

Shattuck-Eidens, D. M., R. N. Bell, S. L. Neuhausen and T. Helentjaris, 1990 DNA sequence variation within maize and melon: observations from polymerase chain reaction amplification and direct sequencing. Genetics **126:** 207–217.

Tajima, F., 1989 Statistical methods for testing the neutral mutation hypothesis by DNA polymorphism. Genetics **123:** 585–595.

Veit, B., R. J. Schmidt, S. Hake and M. F. Yanofsky, 1993 Maize floral development: new genes and old mutants. Plant Cell **5:** 1205–1215.

Veit, B., S. P. Briggs, R. J. Schmidt, M. F. Yanofsky and S. Hake, 1998 Regulation of leaf initiation by the *terminal ear1* gene of maize. Nature **393:** 166–168.

Wang, R.-L., A. Stec, J. Hey, L. Lukens and J. Doebley, 1999 The limits of selection during maize domestication. Nature **398:** 236–239.

Watterson, G. A., 1975 On the number of segregating sites in genetical models without recombination. Theor. Popul. Biol. **7:** 188–193.

Wolfe, K. H., P. M. Sharp and W.-H. Li, 1989 Rates of synonymous substitution in plant genes. J. Mol. Evol. **29:** 208–211.

Communicating editor: A. G. Clark