Nucleotide Polymorphism in the *Adh1* Locus of Pearl Millet (*Pennisetum glaucum*) (Poaceae)

Brandon S. Gaut*,† and Michael T. Clegg*

**Department of Botany and Plant Sciences, University of California, Riverside, Riverside, California 92521, and †Program in Statistical Genetics, Department of Statistics, North Carolina State University, Raleigh, North Carolina 27695*

Manuscript received June 4, 1993
Accepted for publication August 25, 1993

ABSTRACT

We investigated nucleotide polymorphism in the *Adh1* locus of pearl millet (*Pennisetum glaucum*) (Poaceae) by determining the DNA sequence of 20 alleles from 10 individuals. The individuals were sampled from throughout pearl millet's indigenous range and represent both wild and cultivated accessions. Our results indicated that there is little nucleotide polymorphism in the *Adh1* locus. Estimates of per site nucleotide polymorphism did not differ significantly between cultivated and wild millet accessions. We compared nucleotide polymorphism in pearl millet *Adh1* with nucleotide polymorphism in maize (*Zea mays*) *Adh1* and conclude that the maize *Adh1* sample is more polymorphic. Increased polymorphism in maize *Adh1* may be attributable, in part, to faster substitution rates in the maize lineage. Analysis suggests that substitution rates in the maize *Adh1* lineage are ~1.7 times faster than substitution rates in the millet *Adh1* lineage.

POPULATION geneticists traditionally have sought to describe genetic variation and the evolutionary forces shaping genetic variation. Variation can be assessed by isozyme electrophoresis, and differences in allozyme frequencies between populations have been used to infer the evolutionary forces of selection, drift and migration. However, isozymes detect only a subset of nonsynonymous nucleotide substitutions and thus can underestimate genetic variation. This fact has led to the development of more sensitive DNA-based techniques for detecting genetic variation. For example, restriction fragment length polymorphisms (RFLPs) have been used extensively to describe genetic variation. More recently, allelic samples of nucleotide sequences have provided estimates of genetic variation (*e.g.*, KREITMAN 1983). A plethora of statistical tests have been formulated to examine DNA sequence data for evidence of natural selection, migration and geographic subdivision (SLATKIN and MADDISON 1989, 1990; KREITMAN and HUDSON 1991; MCDONALD and KREITMAN 1991; HUDSON, BOOS and KAPLAN 1992).

Sampling of allelic nucleotide sequences has become common place in *Drosophila* species (KREITMAN 1983; KREITMAN and HUDSON 1991; SCHAEFFER and MILLER 1991; RILEY, KAPLAN and VEUILLE 1992; KLIMAN and HEY 1993; LONG and LANGLEY 1993), but this approach has been applied sparingly to problems in plant evolutionary genetics. To date, this approach has only been used in the assessment of genetic variation in *Zea mays* (maize) and its close relatives (SHATTUCK-EIDENS *et al.* 1990; MACRAE and CLEGG 1992; GAUT and CLEGG 1993; GOLOUBINOFF, PÄÄBO and WILSON 1993).

In this article we present a study of nucleotide polymorphism in the *Adh1* locus of *Pennisetum glaucum* (millet, pearl millet) (Poaceae). Cultivated millet (*P. glaucum* subsp. *glaucum*) presumably originated from a wild progenitor (*P. glaucum* subsp. *monodii*) found in the Sahel region of Africa. Both cultivated and wild millet are primarily outcrossing, and they hybridize freely to one another. Isozyme surveys suggest that cultivated millet is as genetically diverse as wild millet (TOSTAIN, RIANDEY and MARCHAIS 1987). This observation has been used to support hypotheses suggesting multiple centers of pearl millet domestication (TOSTAIN, RIANDEY and MARCHAIS 1987) although alternative hypotheses espousing a single center of domestication cannot be dismissed entirely (TOSTAIN 1992). Isozyme surveys also suggest that wild millet is geographically subdivided (TOSTAIN, RIANDEY and MARCHAIS 1987; TOSTAIN 1992).

The *Adh1* gene is a single-copy nuclear gene encoding alcohol dehydrogenase (alcohol NAD+: oxidoreductase, E.C. 1.1.1.1), a protein important in plant response to anoxia (FREELING and BENNETT 1985). The millet *Adh1* gene product has been analyzed extensively by isozyme electrophoresis. The *Adh1* locus is highly diverse in both wild and cultivated millet accessions and is geographically subdivided with respect to allozyme frequency (TOSTAIN 1992; TOSTAIN, RIANDEY and MARCHAIS 1987). RFLP studies also suggest a good deal of variation at the *Adh1* locus (GEPTS and CLEGG 1989). RFLP variation at the *Adh1* locus: (i) is fairly equally distributed between wild and

TABLE 1

List of individuals by location sampled, identification number, status (wild or cultivated) and their *Adh1* haplotype

Country	Individual	Cult./Wild	Haplotypes
Mauritania	85-27	Cult.	*8527A, 8527B*
Tunisia	85-31	Cult.	*8531, 8531*
Gambia	85-15	Cult.	*8514, 8531*
Gambia	85-14	Cult.	*8514,8531*
Niger	85-6	Wild	*856A, 856B*
Senegal	85-34	Wild	*8534A, 8534B*
Senegal	85-35	Wild	*8535, 8535*
Mali	85-37	Wild	*8537, 8537*
Mali	85-38	Wild	*8531, 8535*
Sudan	85-2	Wild	*852A, 852B*
	Tift23DB[a]	Cult.	*23DB*

21 total alleles; 13 distinct haplotypes.
[a] GAUT and CLEGG 1991.

domesticated millets and (ii) tends to be geographically distributed (GEPTS and CLEGG 1989).

It is of interest to compare millet with maize (*Zea mays*) (Poaceae). Both plants are annual grasses, both plants are primarily outcrossing, and both plants have been characterized by extensive isozyme sampling. At the isozyme level, millet *Adh1* may be more diverse than maize *Adh1*. For example, *Adh1* allozyme diversities for populations of wild millet throughout Africa range from 0.35–0.53 (TOSTAIN 1992) while Guatemalan races of maize have *Adh1* allozyme diversities ranging from 0.00–0.43 with an average diversity of 0.04 (BRETTING, GOODMAN and STUBER 1990). However, both RFLP and sequencing analyses of maize *Adh1* reveal extensive variation at the *Adh1* locus of maize (JOHNS, STROMMER and FREELING 1983; ZENG 1992; GAUT and CLEGG 1993).

We sequenced ~2 kb segments from each of 20 *Adh1* alleles from pearl millet and asked three specific questions: (i) how variable is the *Adh1* locus at the nucleotide level? (ii) how does *Adh1* nucleotide polymorphism in millet compare to *Adh1* nucleotide polymorphism in maize? and (iii) can we infer the evolutionary forces contributing to nucleotide polymorphism in the *Adh1* locus?

MATERIALS AND METHODS

We sequenced 1985 bp of the *Adh1* locus from 10 individuals (Table 1). Four of the 10 individuals represent cultivated millet (*P. glaucum* subsp. *glaucum*) and six individuals represent the presumed wild progenitor (*P. glaucum* subsp. *monodii*). The 10 individuals were sampled from a wide geographic distribution ranging from the Atlantic coast of Africa to the Sudan. A single individual was sampled from Tunisia in Northern Africa.

Nucleotide sequences were determined as follows. DNA was isolated from individuals (GEPTS and CLEGG 1989) and *Adh1* nucleotide sequences were amplified using the polymerase chain reaction (PCR). The primers for amplification were based on the *Adh1* sequence from pearl millet (GAUT and CLEGG 1991) and were designed specific to sequences

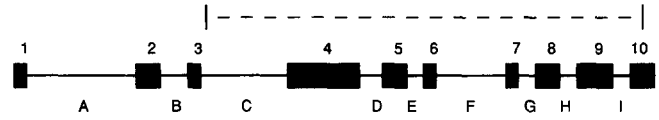

FIGURE 1.—A schematic representation of the *Adh1* gene. Exons are represented by darkened blocks and are labeled from 1 → 10. Introns are solid lines labeled A → I. The coding region spans ~2750 bp. The dashed line represents the region sequenced for all alleles in this study; this region spans ~2000 bp.

in the third and 10th exons. PCR amplification generated a fragment that includes most of the 3′ portion of the *Adh1* coding region (Figure 1). This region corresponds to nucleotide positions 1262 to 3223 (as listed in Figure 1 of GAUT and CLEGG 1991) and includes ~880 bp of exon sequence. This region is homologous to the sequences analyzed in a study of *Adh1* nucleotide polymorphism in the genus Zea (GAUT and CLEGG 1993).

To limit *Taq* polymerase artifacts, we sequenced PCR products directly (SANGER, NICKLEN and COULSON 1977). Because PCR amplification of diploid individuals results in the amplification of two alleles, direct sequencing of PCR products resulted in multiple bands at some nucleotide positions. These multiple bands reflected polymorphisms between *Adh1* alleles in heterozygous individuals. To assign polymorphisms to an allele, we applied the algorithm of CLARK (1990). All alleles except those from three individuals (inds. 85-6, 85-34 and 85-2) were resolved by this method. Alleles from these individuals were resolved by a combination of molecular approaches. In some cases polymorphic sites were near an indel. If the individual was heterozygous for the indel, it was possible to determine linkage of the indel and a polymorphic site by careful sequencing of both DNA strands. For example, the linkage between sites 2601 and the indel at site 2611 in individual 85-34 was resolved by direct sequencing of PCR products. This method works only for polymorphic sites near (*i.e.* within ~100 bp) of a heterozygous indel, but, coupled with the algorithm of CLARK (1990), led to the resolution of alleles *8534A* and *8534B*. Alleles from individuals 85-2 and 85-6 were resolved by sequencing clones of PCR products. In total, our sample of millet alleles is 21: two alleles from each of 10 individuals and another allele from published data (GAUT and CLEGG 1991).

Pearl millet *Adh1* sequences have been deposited in GenBank (accession nos. L20575 → L20586).

Assuming that a sample is representative of a random sample from a panmictic population, we can estimate per site nucleotide polymorphism by $\hat{\theta}$, where $\hat{\theta} = S/(a_{n-1}m)$. In this expression n is the number of genes in the sample, S is the number of polymorphic silent sites in the sample, m is the number of sites in the sample and a_{n-1} is given by $\sum_{i=1}^{n-1} 1/i$ (WATTERSON 1975). Under the infinite sites model with neutral mutations, $\hat{\theta}$ estimates $\theta = 4N\mu$, where N is the population size and μ is the mutation rate. A 95% confidence interval for θ was computed (KREITMAN and HUDSON 1991). Silent sites are defined as intron sites and sites at the third position of a codon.

RESULTS

Sequence data and nucleotide polymorphism: Table 1 lists the individuals sampled as well as the haplotypes found in each individual. Haplotypes were named for the individual in which they were first found. Of the 10 individuals, three were homozygous

(85-31, 85-35 and 85-27) and two (85-15 and 85-14) were identical heterozygotes. Some haplotypes were found in multiple individuals. For example, the *8531* haplotype was found in the cultivated individual from Tunisia (individual 85-31), in two cultivated individuals from Gambia (individuals 85-14 and 85-15) and in a wild individual from Mali (individual 85-38).

The matrix of polymorphic sites is reported in Table 2. There are 23 substitutions in the sample of 21 alleles. Of these, 16 are in introns, five are synonymous substitutions at third codon positions, and two result in amino acid replacements. Each nonsynonymous substitution is found in two haplotypes from a single individual. For example, haplotypes *852A* and *852B* from the Sudanese individual encode an Asp ↔ Asn replacement in residue 89 (nucleotide site 1825 in Table 2), and haplotypes *8527A* and *8527B* from the Mauritania individual encode an Asp ↔ Val replacement in residue 121 (nucleotide site 1922 in Table 2). Both substitutions cause a change in the charge of the amino-acid residue.

There are eight insertion/deletion (indel) polymorphisms in introns (Table 2). There are five indels consisting of a single nucleotide; four of these are unique to a single haplotype. Three indels have lengths of >1 nucleotide (measuring 2, 6 and 15 bp). Each of the larger indels is shared by two or more haplotypes.

The algorithm used to resolve the identity of alleles can underestimate the role of recombination in creating unique haplotypes (CLARK 1990). Despite this bias, haplotype *856B* appears to be a recombinant. For all haplotypes in the sample except *856B*, polymorphic states at sites 1489, 1491, 2375 and 2611 are associated (Table 2). (For example, in haplotypes *8531*, *8534B* and *8537* the deletion at site 2611 is found with the deletion at site 1489, character state "A" at site 1491, and character state "A" at site 2375.) A recombination event somewhere between nucleotide positions 2375 and 2611 adequately accounts for the disassociation of these sites in haplotype *856B*.

Both isozyme and RFLP data reveal comparable levels of *Adh1* variation between cultivated and wild millets (TOSTAIN, RIANDEY and MARCHAIS 1987; GEPTS and CLEGG 1989). Sampling of alleles from cultivated and wild individuals allows comparison of cultivated and wild *Adh1* nucleotide polymorphism. We estimated θ for the entire sample of alleles ($\hat{\theta}_{millet}$), for the sample of alleles found in wild individuals ($\hat{\theta}_{wild}$), and for the sample of alleles found in cultivated individuals ($\hat{\theta}_{cult}$) (Table 3). We compared these three estimates by their 95% confidence intervals and conclude that they do not differ significantly from one another.

Geographic subdivision: Isozyme and RFLP analyses suggest that there is population subdivision with

TABLE 2

Matrix of substitutions between pearl millet haplotypes

	Nucleotide site																															
	1	1	1	1	1	1	1	1	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	3	3
	3	4	4	4	5	5	5	7	8	1	3	3	3	3	4	4	5	5	5	6	6	6	6	7	8	0	4	2	8	1
	7	4	4	8	9	5	6	9	9	4	2	3	6	7	4	6	1	2	8	0	1	4	2	4	1	4	1	4	0b	2
haplo	2	5b	6b	9	1	0	1	7	0	6	2	7	0	5	9	2	6	1	4	6	1	9b	5	3	0	4	0	7	0b	7
23DB	A	X	X	2	C	T	C	G	G	T	G	T	A	A	T	C	C	G	A	T	A	A	T	A	16	X	C	T	T	D	G
8527A	—	T	—	—	—	—	T	—	—	—	—	—	T	—	—	—	—	—	—	—	—	—	—	—	—	T	—	—	—	—	—
8527B	—	T	—	—	—	—	T	—	—	—	—	—	T	—	—	A	—	—	—	—	—	—	—	—	—	T	—	—	—	—	—
8531	—	T	—	D	A	—	T	—	—	C	—	—	—	—	—	—	—	—	G	C	—	D	T	—	—	—	C	C	—	—	—
856B	—	T	—	—	—	—	T	—	—	—	—	—	—	—	—	A	—	—	G	C	—	D	T	—	—	—	—	C	—	6*	—
856A	—	T	—	—	—	—	T	—	—	—	—	—	—	—	—	—	—	—	T	G	—	—	T	—	—	—	C	C	—	—	—
8514	—	T	—	—	—	—	T	C	C	—	—	—	X	—	—	A	—	—	—	G	C	G	T	—	—	—	—	C	—	—	—
8534B	—	T	T	D	A	—	T	C	C	—	—	—	—	—	—	A	—	—	G	C	C	D	T	—	—	—	—	C	—	6*	—
8534A	—	T	—	—	C	T	C	C	C	—	—	—	—	—	—	—	—	—	G	C	C	—	T	—	—	—	—	C	—	—	—
8535	—	T	—	—	—	T	—	C	C	—	—	—	—	G	—	—	—	—	G	C	C	D	T	—	—	—	—	C	—	T	—
8537	—	T	T	D	A	—	T	—	—	—	—	—	—	—	—	A	—	—	G	C	C	—	T	—	—	—	—	C	—	—	—
852A	—	T	—	—	—	—	T	C	C	A	—	—	—	—	—	—	—	—	G	C	C	—	T	—	—	—	—	C	—	—	—
852B	G	T	—	—	—	—	T	C	C	A	—	—	—	—	—	C	—	—	G	C	C	—	T	—	—	—	—	C	—	—	—
pos.										1	3	2				3			3					3			3	3			

* TTCAGT.

Haplotype identities are given in Table 1. Nucleotide sites are numbered according to GAUT and CLEGG (1991). Dashes represent character states identical to the reference *23DB* sequence. Single bp deletions are represented as "X," deletions of >1 bp are represented as "D" and their length is given. The insertion at position 3040b has not been previously reported and is given. The codon position, pos., of polymorphic sites within exons is also provided.

TABLE 3

Estimates of θ

	Alleles	Haplo	S	$\hat{\theta}$	$\hat{\theta}_{0.025}$	$\hat{\theta}_{0.975}$
Silent substitutions						
Millet	21	13	21	0.0042	0.0020	0.0098
Cult	9	5	9	0.0024	0.0009	0.0084
Wild	12	9	15	0.0036	0.0016	0.0102
Maize	6	6	64	0.0189	0.0081	0.0681

Millet estimates are grouped into cultivated (cult), wild or all (millet) data and are based on 1376 silent sites. The maize estimate is based on 1483 silent sites (GAUT and CLEGG, 1993). The number of alleles, the number of haplotypes (haplo), and the number of polymorphic sites (S) are given. $\hat{\theta}_{0.025}$ and $\hat{\theta}_{0.975}$ represent the upper and lower bounds of the 95% confidence intervals of θ.

TABLE 4

TAJIMA (1989) test for neutrality of nucleotide polymorphism

	S	$\hat{M}$	$\hat{k}$	D	*P* value
All alleles: $n = 21$					
All events	31	8.617	6.041	−1.612	>0.10
All subs	23	6.393	4.218	−1.294	>0.10
Indels	8	2.223	1.914	−0.758	>0.10
Wild alleles: $n = 12$					
All events	21	6.954	5.583	−0.869	>0.10
All subs	16	5.298	3.847	−1.182	>0.10
Indels	5	1.656	1.736	0.179	>0.10
Cult. alleles: $n = 9$					
All events	15	5.519	5.580	0.054	>0.10
All subs	10	3.679	3.852	0.218	>0.10
Indels	5	1.840	1.728	−0.256	>0.10

regard to the *Adh1* locus (TOSTAIN 1992; GEPTS and CLEGG 1989). Our sample of alleles was taken from widely divergent geographic locales thought to represent distinct groups of millet. We wished to examine the sample of 21 alleles for evidence of population subdivision.

On average, alleles within a subdivided population will be more closely related to one another than alleles from different subpopulations. If there is geographic subdivision with respect to our sample, the *Adh1* alleles from within individuals will be more closely related to one another than alleles from between individuals. We examined this hypothesis as follows: we calculated the average two-parameter pairwise distance (KIMURA 1980) between alleles within the 10 individuals and compared this average with an average pairwise distance between 10 pairs of alleles sampled with replacement from the sample of 21 *Adh1* alleles. We sampled 1000 times; if the observed average was less than the sampled average 95% of the time, this may have been evidence of subdivision. It was clear that alleles within individuals were more closely related than alleles sampled at random ($P = 0.005$).

Intraspecific tests for selection: TAJIMA (1989) formulated a statistic to test for the neutrality of polymorphisms by comparing two different estimators. The measure $\hat{M}$, where $\hat{M} = \hat{\theta}m$, is not a function of the frequencies of polymorphic states in a sample of alleles. The estimator $\hat{k}$, the average number of pairwise differences between the sequences, does rely on the frequencies of polymorphic sites in a sample (TAJIMA 1983). Under a null hypothesis of neutral nucleotide polymorphisms, these measures of polymorphism ($\hat{M}$ and $\hat{k}$) should not vary significantly (TAJIMA 1989). We applied the TAJIMA test to *P. glaucum Adh1* data. We tested for departures of neutrality in all alleles, in alleles found in cultivated individuals, and in alleles found in wild individuals. In each group of alleles, we examined all polymorphisms (*i.e.*, nucleotide substitutions and indels), only nucleotide substitutions and only indels. We could not reject the null hypothesis in any test (Table 4).

FU and LI (1993) have also formulated a test for the neutrality of polymorphisms. This test statistic relies on the difference between the number of polymorphic sites in external phylogenetic branches (that is, polymorphisms unique to an extant sequence) and number of polymorphic sites in internal phylogenetic branches (that is, polymorphisms shared by extant sequences in the sample). Optimal utilization of this test requires an outgroup to the sample of sequences. There is no clear outgroup in this sample of 21 alleles, however, so we apply the test without outgroup (as described in FU and LI (1993)). For a sample of 21 sequences with 23 polymorphic sites and 14 unique sites, the test statistic, D*, equals −1.754. Given the assumptions of this test (*i.e.*, no migration and constant population size), $D^* < 0.00$ indicates an excess of unique mutations. However, the test is not significant at $\alpha = 0.05$, and thus we cannot reject the null hypothesis of neutral polymorphisms.

Interspecific comparisons of nucleotide polymorphism: Table 3 reports values of $\hat{\theta}$ calculated from maize and millet *Adh1* nucleotide sequences. The estimate of θ for maize is based upon six *Z. mays* sequences (GAUT and CLEGG 1993). The 95% confidence interval of θ_{maize} and θ_{millet} do overlap to some degree (Table 3) and thus we cannot conclude they differ at a significance level of $\alpha = 0.05$. However, 90% confidence intervals for θ_{maize} and θ_{millet} do not overlap (data not shown), and the maize sample clearly contains more nucleotide variation than the millet sample.

Both millet and maize allelic samples have been tested for neutrality of polymorphisms (GAUT and CLEGG 1993 and above). Neither sample of alleles shows evidence of departure from neutrality. However, previous tests did not take advantage of information between species. MCDONALD and KREITMAN (1991) proposed a test to determine whether differ-

TABLE 5

Relative rate tests

Outgroup	$K_{maize} - K_{millet}$	SE	*P* value
Rice	0.0217	0.0067	<0.001
Barley	0.0229	0.0069	<0.001

ences between alleles of different species are due to the accumulation of neutral mutations by random drift or due to fixation of mutations by selection. This test compares the ratio of nonsynonymous to synonymous substitutions for nucleotide differences fixed between species and for nucleotide differences polymorphic within species. Under the null hypothesis of neutral polymorphisms, the ratio of fixed differences will not differ significantly from the ratio of polymorphic differences.

We limit ourselves to exon data for this test because of the difficulty in assigning between species homology for intron sites. In exons 4–10, the six maize *Adh1* alleles have three nonsynonymous polymorphisms and 14 synonymous polymorphisms. Pearl millet alleles have two nonsynonymous polymorphisms and five synonomous polymorphisms (Table 2), and thus the combined ratio of polymorphic substitutions among species is 5:19. The ratio of nonsynonymous to synonymous fixed differences between species is 5:57 (data not shown). A Fisher exact test indicates that these ratios are equivalent at $\alpha = 0.05$ ($P = 0.133$), and thus departures from neutrality are not detected by this method.

Differences in mutation rates and/or selection between maize and millet *Adh1* lineages may be detected by relative rate tests. Previous studies have examined substitution rates in maize and millet *Adh1* lineages and have concluded that maize and millet *Adh1* do not differ in their rates of nucleotide substitution (GAUT and CLEGG 1991). We wished to test this conclusion more thoroughly by utilizing allelic *Adh1* data. LI and BOUSQUET (1992) have formulated a relative rate test for cases where more than one sequence is available from an evolutionary lineage. We apply this test to millet and *Z. mays* haplotype samples using either the rice *Adh1* (XIE and WU 1989) or the barley *Adh1* sequence (GOOD, PELCHER and CROSBY 1988) as an outgroup. The test is performed using two-parameter distances (KIMURA 1980), and we again limit the test to exon regions in order to insure accurate assessment of homology between sequences. The relative rate test yields significant differences in substitution rates between maize and millet *Adh1* lineages (Table 5). This result is dependent upon testing groups of sequences; relative rate tests examining single pairs of maize and millet sequences yield no significant results (data not shown).

DISCUSSION

Polymorphism in millet: Nucleotide data from the *Adh1* locus of pearl millet exhibit little polymorphism. In a sample of 21 alleles, each of ~2000 bp, there are 13 unique haplotypes; one of these haplotypes is a recombinant. The 13 haplotypes contain 23 nucleotide substitutions and 8 indels. Two nucleotide substitutions are nonsynonymous; both of these code for amino acid replacements that change the charge of the protein. Isozyme electrophoresis reveals that the alleles with nonsynonymous substitutions are electrophoretic variants (data not shown). Although it cannot be determined that changes in electrophoretic mobility are due to these particular nonsynonymous substitutions (because our nucleotide data lack some portions of the *Adh1* coding region), isozyme data are consistent with the hypothesis that charge-changing amino-acid replacements confer changes in allozyme mobility.

There are two primary pearl millet domestication hypotheses: the single domestication center hypothesis (BRUNKEN, DEWET and HARLAN 1977) and the multiple center hypothesis (HARLAN 1971). Under the former hypothesis one might expect a genetic bottleneck to reduce genetic variation in cultivars. Neither isozyme data nor *Adh1* RFLP data detect such a reduction (TOSTAIN, RIANDEY and MARCHAIS 1987; GEPTS and CLEGG 1989). Nucleotide data are consistent with these observations; there is no detectable reduction in nucleotide polymorphism at the *Adh1* locus of cultivated individuals. However, rDNA in cultivated millet accessions appears to be less variable than rDNA in wild millet accessions, suggesting a reduction of rDNA variation in cultivars (GEPTS and CLEGG 1989). Given that free hybridization between cultivated and wild millet may obscure patterns of genetic diversity, our data provide no new insights into the process of pearl millet domestication.

Previous work has documented geographic subdivision of *Adh1* allozymes (TOSTAIN 1992; TOSTAIN, RIANDEY and MARCHAIS 1987) and RFLP haplotypes (GEPTS and CLEGG 1989). Our results suggested that alleles in our sample were more closely related within individuals than among individuals. While this observation may reflect the existence of selfing (millet is primarily an outcrossing plant but selfing can occur (BURTON 1974)) and/or the existence of local population subdivision, it is consistent with previous results documenting geographic subdivision of *Adh1* alleles.

Interspecific comparisons: There was more nucleotide variation in our sample of maize *Adh1* alleles than in our sample of millet *Adh1* alleles. This is reflected in differences in $\hat{\theta}_{maize}$ and $\hat{\theta}_{millet}$ (Table 3). There may be many reasons for observed differences in $\hat{\theta}$. First, this result could be an artifact of sampling. Few alleles were sampled in maize, and these were not

sampled randomly; the *Adh1-C*^{m} (OSTERMAN and DENNIS 1989) and the *Adh1-S* (DENNIS *et al.* 1984) alleles were sequenced because they encode different allozymes. Surprisingly, the two nonrandomly sampled alleles contribute no new silent polymorphism to the maize sample, and consequently $\hat{\theta}$ is greater in the sample of four presumably random alleles ($\hat{\theta} = 0.023$) than in the total sample of six maize alleles. Thus, it does not appear that nonrandom sampling of alleles is responsible for greater levels of nucleotide polymorphism in maize. However, this may indicate that the small sample size of *Z. mays* alleles is problematic.

Second, it is assumed that allelic samples are taken from an equilibrium population. Both maize and millet are geographically differentiated (TOSTAIN 1992; BRETTING, GOODMAN and STUBER 1990) and hence neither allelic sample is from a single equilibrium population. Selection at or near the *Adh1* locus may also violate the equilibrium assumption. It must be noted, however, that violations of equilibrium assumptions may not be severe. Neither intraspecific nor interspecific tests detect deviation from neutrality. Failure to detect a deviation suggests either that departures from the equilibrium neutral model are not great or that the tests have low power to reject neutrality. If departures from the neutral model are small, differences in $\hat{\theta}$ may reflect differences in population size and mutation rates (see below).

Unequal substitution rates between *Adh1* lineages may be a function of unequal mutation rates or differential selection. Differential selection can account for heterogeneous substitution rates if: (i) the accelerated lineage has a historically smaller population size, (ii) the accelerated lineage has undergone repeated fixations of advantageous alleles or (iii) purifying selection is less stringent in the accelerated lineage. In the first case, a smaller population size will result in a higher proportion of mutations being neutral rather than selectively deleterious because neutrality is defined as $NS \ll 1$ (KIMURA 1983), where N is the population size and S is the selection coefficient. In the second case (repeated fixations of an advantageous allele), positive selection will limit genetic diversity. In both of these cases, one would expect nucleotide polymorphism to be reduced in the accelerated lineage. This is not observed for *Adh1* sequence data, and thus neither of these two possibilities is consistent with the data.

In the case of differential purifying (or constraining) selection between lineages, one would expect more nucleotide polymorphism in the accelerated lineage as a result of reduced structural constraint on the protein. We find more polymorphism in the accelerated lineage and thus this hypothesis may be consistent with *Adh1* nucleotide data. However, this view is inconsistent with maize and millet allozyme data. One would expect differential purifying selection to affect primarily the amount of phenotypic variability at a locus. Assuming that allozyme diversity is a good measure of phenotypic variability (*i.e.*, variation in enzyme kinetic parameters), millet *Adh1* may exhibit more phenotypic variability than maize *Adh1*. This contradicts the view that purifying selection is reduced in the maize lineage relative to the millet lineage. Furthermore, one would expect differential purifying selection between lineages to minimally impact silent sites, but silent site differences are the largest component of the distances used in relative rate tests. Thus, it does not appear that differences in purifying selection between lineages adequately accounts for accelerated substitution rates in maize.

TABLE 6

Estimating the ratio of substitution rates

Outgroup	Maize:millet	95% C.I.
Rice	1.70	1.60, 1.80
Barley	1.77	1.67, 1.87

Maize:millet represents the difference in substitution rates between lineages. 95% C.I. represents the confidence interval of the ratio estimate.

Alternatively, differences in substitution rates can be fueled by differences in mutation rates. If this is indeed the case, it is desirable to estimate the difference in mutation rates between maize and millet *Adh1* lineages. GAUT *et al.* (1992) estimated differences in substitution rates between evolutionary lineages using the maximum likelihood framework of MUSE and WEIR (1992). Application of these techniques suggest that maize *Adh1* has a substitution rate ~1.7 times faster than millet *Adh1* (Table 6). Assuming that differences in substitution rates are primarily attributable to heterogeneous mutation rates, we postulate that mutation occurs ~1.7 times more frequently in maize *Adh1* relative to millet *Adh1*.

Assuming the infinite alleles model and neutrality of polymorphic silent sites, $\hat{\theta}$ estimates $4N\mu$. If variation in $\hat{\theta}$s reflect differences in N and μ between species, one can use $\hat{\theta}_{maize}$ and $\hat{\theta}_{millet}$ to compare the historical population sizes of the two grass species. This calculation is complicated by potentially different mutation rates (μ) in the two *Adh1* lineages. Given that θ estimates are 4.5-fold different between maize and millet lineages (Table 3) and that mutation rates may be ~1.7-fold different between maize and millet lineages (Table 6), we estimate that maize has a historical effective population size ~2.6-fold larger than millet. This result holds only if the assumptions of the equilibrium neutral model are not seriously violated.

Allozyme diversity in maize *Adh1* is no greater than allozyme diversity in millet *Adh1* yet maize is apparently more variable on the nucleotide level. Any of a

number of evolutionary forces (*e.g.*, drift, selection, migration) could produce these discordant observations, but at present there is no way to distinguish the relative contribution of each evolutionary process. More thorough sampling of maize and millet alleles may allow better insight into the paradox of relatively high allozyme diversity but relatively low nucleotide polymorphism in the *Adh1* locus of pearl millet.

We would like to thank J. CLEGG for isozyme analyses, B. S. WEIR for helpful discussion and J. F. DOEBLEY, A. G. CLARK and an anonymous reviewer for helpful comments. R. R. HUDSON suggested tests for subdivision. This study was supported by National Institutes of Health grant GM45155 to M.T.C., National Institutes of Health Postdoctoral Fellowship GM15528 to B.S.G., and National Institutes of Health grant GM45344 to North Carolina State University.

LITERATURE CITED

BRETTING, P. K., M. M. GOODMAN and C. W. STUBER, 1990 Isozymatic variation in Guatemalan races of maize. Am. J. Bot. **77:** 211–225.

BRUNKEN, J., J. M. J. DEWET and J. R. HARLAN, 1977 The morpholgy and domestication of pearl millet. Econ. Bot. **31:** 163–174.

BURTON, G. W., 1974 Factors affecting pollen movement and natural crossing in pearl millet. Crop Sci. **14:** 802–805.

CLARK, A. G., 1990 Inference of haplotypes from PCR-amplified samples of diploid populations. Mol. Biol. Evol. **7:** 111–122.

DENNIS, E. S, W. L. GERLACH, A. J. PRYOR, J. L. BENNETZEN, A. INGLIS, *et al.*, 1984 Molecular analysis of the *Alcohol dehydrogenase* (*Adh1*) gene of maize. Nucleic Acids Res. **12:** 3983–4000.

FREELING, M., and D. C. BENNETT, 1985 Maize *Adh1*. Annu. Rev. Genetics **19:** 297–323.

FU, Y.-X., and LI, W.-H., 1993 Statistical tests of neutrality of mutations. Genetics **133:** 693–709.

GAUT B. S., and M. T. CLEGG, 1991 Molecular evolution of alcohol dehydrogenase 1 in members of the grass family. Proc. Natl. Acad. Sci. USA **88:** 2060–2064.

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

GAUT, B. S., S. V. MUSE, W. D. CLARK and M. T. CLEGG, 1992 Relative rates of nucleotide substitution at the *rbc*L locus of monocotyledonous plants. J. Mol. Evol. **35:** 292–303.

GEPTS, P., and M. T. CLEGG, 1989 Genetic diversity in pearl millet (*Pennisetum glaucum* [L.]R.Br.) at the DNA sequence level. J. Hered. **80:** 203–208.

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.

GOOD, A., L. E. PELCHER and W. L. CROSBY, 1988 Nucleotide sequence of a complete barley alcohol dehydrogenase 1 cDNA. Nucleic Acids Res. **16:** 7182.

HARLAN, J., 1971 Agricultural origins: centers and non-centers. Science **14:** 1–44.

HUDSON, R. R., D. D. BOOS and N. L. KAPLAN, 1992 A statistical test for detecting geographic subdivision. Mol. Biol. Evol. **9:** 138–151.

JOHNS, M. A., J. N. STOMMER and M. FREELING, 1983 Exceptionally high levels of restriction site polymorphism in DNA near the maize *Adh1* gene. Genetics **105:** 733–743.

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.

KIMURA, M., 1983 *The Neutral Theory of Molecular Evoltuion*. Cambridge University Press, Cambridge, p. 44.

KLIMAN, R. M., and J. HEY, 1993 DNA sequence variation at the *period* locus within and among species of the *Drosophila melanogaster* complex. Genetics **133:** 375–387.

KREITMAN, M., 1983 Nucleotide polymorphsim at the alcohol dehydrogenase locus of *Drosophila melanogaster*. Science **304:** 412–417.

KREITMAN, M., and R. R. HUDSON, 1991 Inferring the evolutionary histories of the *ADh* and *Adh-dup* loci in *Drosophila melanogaster* from patterns of polymorphism and divergence. Genetics **127:** 565–582.

LI, P., and J. BOUSQUET, 1992 Relative-rate test for nucleotide substitution between two lineages. Mol. Biol. Evol. **9:** 1185–1189.

LONG, M., and C. H. LANGLEY, 1993 Natural selection and the origin of *jingwei*, a chimeric processed function gene in *Drosophila*. Science **260:** 91–95.

MACRAE, A., and M. T. CLEGG, 1992 Evolution of *Ac* and *Ds1* elements in select grasses (Poaceae). Genetics **86:** 55–66.

MCDONALD, J. H., and M. KREITMAN, 1991 Adaptive protein evolution at the *Adh1* locus in *Drosophila*. Nature **351:** 652–654.

MUSE, S. V., and B. S. WEIR, 1992 Testing for equality of evolutionary rates. Genetics **132:** 269–276.

OSTERMAN, J. C., and E. S. DENNIS, 1989 Molecular analysis of the *Adh1-C*^{m} allele of maize. Plant Mol. Biol. **13:** 203–212.

RILEY, M. A., S. R. KAPLAN and M. VEUILLE, 1992 Nucleotide polymorphism at the Xanthine Dehydrogenase locus in *Drosophila pseudoobscura*. Mol. Biol. Evol. **9:** 56–69.

SANGER, F. S., S. NICKLEN and A. R. COLSON, 1977 DNA sequencing with chain terminating inhibitors. Proc. Natl. Acad. Sci. USA **74:** 5463–5467.

SCHAEFFER, S. W., and E. L. MILLER, 1991 Nucleotide sequence analysis of *Adh* genes estimates the time of geographic isolation of the Bogota population of *Drosophila pseudoobscura* Proc. Natl. Acad. Sci. USA **88:** 6097–6101.

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.

SLATKIN, M., and W. P. MADDISON, 1989 A cladistic measure of gene flow inferred from the phylogenies of alleles. Genetics **123:** 603–613.

SLATKIN, M., and W. P. MADDISON, 1990 Detecting isolation by distance using phylogenies of genes. Genetics **126:** 249–260.

TAJIMA, F., 1983 Evolutionary relationship of DNA sequences in finite populations. Genetics **105:** 437–460.

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

TOSTAIN, S., 1992 Enzyme diversity in pearl millet (*Pennisetum glaucum* L.) 3. Wild millet. Theor. Appl. Genet. **83:** 733–742.

TOSTAIN, S., M. F. RIANDEY and L. MARCHAIS, 1987 Enzyme diversity in pearl millet (*Pennisetum glaucum*). Theor. Appl. Gene. **74:** 188–193.

WATTERSON, G. A., 1975 On the number of segregating sites in genetical models without recombination. Theor. Popul. Biol. **7:** 256–276.

XIE, Y., and R. WU., 1989 Rice alcohol dehydrogenase genes: anaerobic induction, organ specific expression and characterization of cDNA clones. Plant Mol. Biol. **13:** 53–68.

ZENG, A., 1992 Molecular evolution of the *Adh1* gene in maize. Ph.D. Dissertation, Washington University, St. Louis.

Communicating editor: A. G. CLARK