Low Levels of Nucleotide Diversity at Homoeologous *Adh* Loci in Allotetraploid Cotton (*Gossypium* L.)

Randall L. Small, Julie A. Ryburn, and Jonathan F. Wendel

Department of Botany, Iowa State University

Levels of genetic diversity within and among populations and species are shaped by both external (population-level) and internal (genomic and genic) evolutionary forces. To address the effect of internal pressures, we estimated nucleotide diversity for a pair of homoeologous *Adh* loci in an allotetraploid species, *Gossypium hirsutum.* These data represent the first such estimates for a pair of homoeologous nuclear loci in plants. Estimates of nucleotide diversity for *AdhA* in *Gossypium* are lower than those for any plant nuclear gene yet described. This low diversity appears to reflect primarily a history of repeated, severe genetic bottlenecks associated with both speciation and recent domestication, supplemented by an unusually slow nucleotide substitution rate and an autogamous breeding system. While not statistically supportable, the sum of the observations also suggest differential evolutionary dynamics at each of the homoeologous loci.

Introduction

Levels and patterns of genetic diversity vary greatly within and among populations and species. This variation reflects the interplay of myriad historical factors and evolutionary forces, involving external forces such as natural selection, population size and history, gene flow, and breeding system, as well as internal genomic and genetic factors such as recombination, mutation rate, and gene conversion (Aquadro and Begun 1993; Tajima 1993*b*; Moriyama and Powell 1996; Clegg 1997; Clegg, Cummings, and Durbin 1997; Amos and Harwood 1998). Recent studies have revealed varying patterns of nucleotide diversity within plant species (Gaut and Clegg 1993*a*, 1993*b*; Hanfstingl et al. 1994; Hanson et al. 1996; Innan et al. 1996; Miyashita, Innan, and Terauchi 1996; Huttley et al. 1997; Kawabe et al. 1997; Terauchi, Terachi, and Miyashita 1997; Bergelson et al. 1998; Cummings and Clegg 1998; Eyre-Walker et al. 1998; Liu, Zhang, and Charlesworth 1998). While these and other studies have yielded a number of insights into the factors that shape naturally occurring variation, in any particular case, the evolutionary or historical forces responsible for the diversity patterns observed may be difficult to discern. This is especially true for comparisons between species, for which numerous potentially confounding life history features and population histories may influence both the amount and apportionment of diversity. Allopolyploid species, which contain duplicated genes in the same nucleus, may be particularly useful in isolating potentially relevant internal genetic and genomic factors from external population-level processes. In allopolyploids, some external processes (e.g., selection) can differentially effect duplicated genes, while others (e.g., genetic drift, breeding system) are expected to effect duplicated genes equivalently. Thus, examining molecular evolution at duplicated genes in a polyploid allows at least some population-level effects to be ruled out as having contributed to differential evolution.

The cotton genus (*Gossypium*) provides a model system for studying molecular evolution of genes duplicated by allopolyploidy. The five tetraploid *Gossypium* species ($n = 26$) are a monophyletic assemblage derived from a single allopolyploidization event that occurred approximately 1–2 MYA (Wendel 1989; Seelanan, Schnabel, and Wendel 1997; Small et al. 1998). A robust phylogenetic framework has been developed for both the diploid and allopolyploid members of the genus (fig. 1; Wendel and Albert 1992; Seelanan, Schnabel, and Wendel 1997; Small et al. 1998). Diploid *Gossypium* species (all $n = 13$) have been divided into genomic groups (A–K) based on differences in chromosome size and pairing behavior in interspecific hybrids (Endrizzi, Turcotte, and Kohel 1985; Stewart 1995). The two diploid species that gave rise to the allotetraploids were from the A-genome and D-genome groups and are best represented by the extant species *G. herbaceum* L. and *G. raimondii* Ulbr., respectively (Endrizzi, Turcotte, and Kohel 1985; Wendel, Schnabel, and Seelanan 1995; Small et al. 1998). Tetraploid species are therefore termed the AD-genome group, and their two constituent genomes are referred to as the A- and D-subgenomes.

Two of the allotetraploid species, *G. hirsutum* L. and *G. barbadense* L., were independently domesticated within the last 5,000 years for their seed fiber (reviewed in Wendel 1995). The genetic consequences of domestication of *G. hirsutum,* the species that presently dominates world cotton commerce, have been explored in depth at the isozyme and restriction fragment length polymorphism (RFLP) levels (Wendel, Brubaker, and Percival 1992; Brubaker and Wendel 1994). Among the conclusions of these studies is that genetic diversity in *G. hirsutum* is very low and is especially restricted in the gene pool represented by modern annualized cultivars. *Gossypium hirsutum* was probably first domesticated in the Yucatan peninsula. The only extant form of *G. hirsutum* that arguably is wild, race ''yucatanense,'' is found here as a common component of the indigenous beach strand vegetation (Stephens 1958; Sauer 1967), where it exists as a sprawling, perennial shrub. Evidence suggests that following initial domestication, the origi-

Key words: alcohol dehydrogenase, cotton, *Gossypium,* homoeology, nucleotide diversity, polyploidy.

Address for correspondence and reprints: Jonathan F. Wendel, Department of Botany, Iowa State University, Ames, Iowa 50011. E-mail: jfw@iastate.edu.

Mol. Biol. Evol. 16(4):491–501. 1999

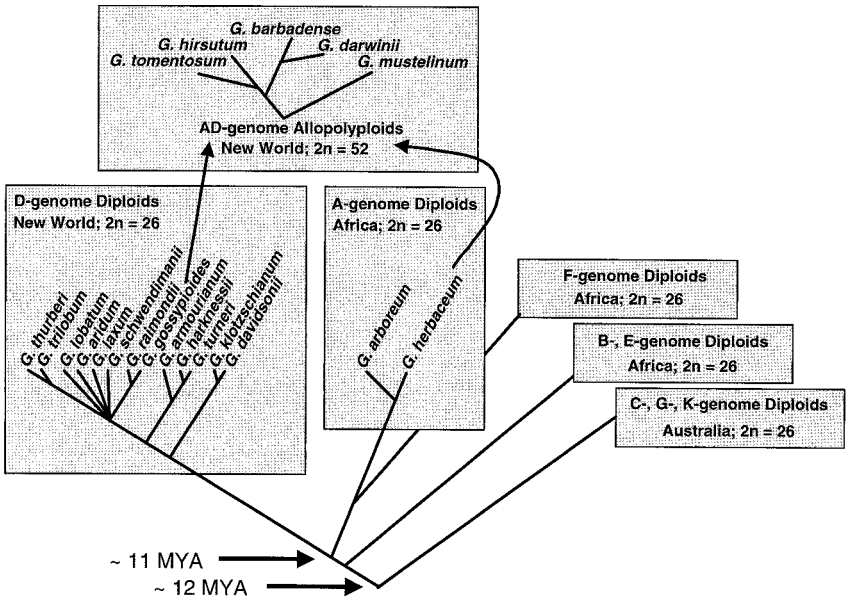

FIG. 1.—Composite phylogenetic hypothesis for the genus *Gossypium* based on analyses of multiple molecular data sets. Individual genome groups are boxed, and their geographic origins and chromosome numbers are identified. This hypothesis depicts both the divergent evolution of the diploid species (A–K genomes; $2n = 26$) and the reticulate origin of the allotetraploids (AD-genome; $2n = 52$) via hybridization and polyploidization between A- and D-genome diploids (best represented by *G. herbaceum* and *G. raimondii*, respectively). Arrows indicate estimated time of divergence for two nodes: the initial divergence within the genus at ca. 12 MYA and the divergence between the A- and D-genome groups at ca. 11 MYA.

nal perennial cultivated forms became widely dispersed throughout the Yucatan peninsula. Later, localized derivatives developed, the most important of which was the annualized race ''latifolium,'' which is suggested to have spread to Guatemala and southern Mexico, where further agronomic development took place, leading to cultivated forms that spread via human-mediated diffusion throughout Mesoamerica. Molecular-marker evidence (Wendel, Brubaker, and Percival 1992; Brubaker and Wendel 1994) shows that most of the gene pool of the modern annual forms of *G. hirsutum*, including the Upland cotton cultivars that are common in the cotton belt of the United States, traces to Mexican stocks that had been transported there from Guatemala and southern Mexico (Niles and Feaster 1984). This history of sequential genetic bottlenecks and rapid population expansion is thought to be responsible for the constrained levels of genetic diversity that are often observed in crop plant gene pools (Doebley 1989, 1992).

The goal of the present study was to quantify nucleotide diversity at a pair of homoeologous *Adh* loci in the allotetraploid species *G. hirsutum.* With these data, we asked the following questions: (1) Are estimates of nucleotide diversity equivalent at the two homoeologous loci that resulted from the allopolyploidization event? (2) How do estimates of nucleotide diversity for *Adh* in cotton compare with nucleotide diversity estimates for other species? (3) Are nucleotide diversity data consistent with previous information on the history of domestication for *G. hirsutum*? (4) How do direct measurements of genetic diversity based on DNA sequence data compare with indirect estimates derived from isozyme and RFLP data?

Materials and Methods

Plant Materials

Based on variation at 205 anonymous nuclear loci detected by low-copy RFLP probes, Brubaker and Wendel (1994) defined 18 genetic/geographic groups of wild or feral *G. hirsutum* that represent the diversity encompassed by the species. In addition, they sampled several cultivars (''Upland cotton'') from the cotton belt of the United States and several accessions of *G. barbadense* (''Pima cotton,'' ''Egyptian cotton''). For this study, we surveyed one accession from each of the 18 groups identified by Brubaker and Wendel (1994) and three Upland cultivars of *G. hirsutum.* We also included five accessions of *G. barbadense* for comparison (table 1). Because one of our objectives was to compare diversity detected using various molecular tools, the same accessions as in earlier studies employing isozymes (Wendel, Brubaker, and Percival 1992) and RFLPs (Brubaker and Wendel 1994) were used. Genomic DNAs used in the present study were the same as those described in Brubaker and Wendel (1994).

PCR Amplification, Cloning, and Sequencing

The *Adh* gene family in *Gossypium* consists of a minimum of five genetic loci in the diploid species and five pairs of homoeologous loci in the allotetraploids (unpublished data). *Adh* genes encode alcohol dehydrogenase (alcohol : NAD^+ oxidoreductase, E.C. 1.1.1.1), metabolic enzymes involved primarily in anaerobic respiration. We have termed the locus discussed here *AdhA. AdhA* exists as a single-copy locus in the diploids *G. herbaceum* (A-genome) and *G. raimondii* (D-genome) and at a pair of homoeologous loci in the allotetraploids

Table 1
Plant Materials Used in this Study

Species	Accession	Geographic Origin
Gossypium hirsutum.	pfx	Oaxaca, Mexico
	TX-1	Guerrero, Mexico
	TX-6	Pueblo, Mexico
	TX-21	Chiapas, Mexico
	TX-44	Chiapas, Mexico
	TX-51	Chiapas, Mexico
	TX-93	Jutiapa, Guatemala
	TX-94	Zacapa, Guatemala
	TX-98	Chiquimula, Guatemala
	TX-111	Jutiapa, Guatemala
	TX-116	Santa Rosa, Guatemala
	TX-119	El Salvador
	TX-166	Zacapa, Guatemala
	TX-188	Baja Verapaz, Guatemala
	TX-192	Oaxaca, Mexico
	TX-367	Santa Rosa, Guatemala
	TX-481	Yucatan, Mexico
	TX-706	Honduras
	TX-766	Belize
	Paymaster H86048	Cultivar
	Deltapine 50	Cultivar
	BR115	Cultivar
Gossypium barbadense . .	Pima S5	Cultivar
	B106	Dominican Republic
	B250	Belize
	B444	Colombia
	B559	Venezuela

(fig. 2). As expected from the organismal history, allotetraploid cotton contains a pair of homoeologous loci corresponding to the copies donated by the A- and D-genome ancestors at the time of allopolyploid formation. While most *Gossypium Adh* genes have the classical 10 exon/9 intron *Adh* structure (Millar and Dennis 1996; unpublished data), *AdhA* has lost introns 4 and 7 (fig. 3). Intron loss has been observed in other plant *Adh* loci (Chang and Meyerowitz 1986; Trick et al. 1988; Charlesworth, Liu, and Zhang 1998), as well as in other plant genes (e.g., Drouin and Moniz de Sá 1997; Frugoli et al. 1998), and is presumably accomplished via gene conversion or recombination between an intact gene and a reverse-transcribed cDNA or processed pseudogene (Drouin and Moniz de Sá 1997; Frugoli et al. 1998). A more complete analysis of the structure and evolution of the *Gossypium Adh* gene family will be presented elsewhere.

To isolate *AdhA* sequences, we designed *AdhA*-specific PCR primers homoeologous to regions in exons 2 and 8 (fig. 3; *Adh*x2-1: CTT CAC TGC TTT ATG TCA CAC T; *Adh*x8-1: GGA CGC TCC CTG TAC TCC) and amplified a ~1-kb fragment of *AdhA.* PCR reactions were performed as described (Small et al. 1998). Because *AdhA* exists as a pair of homoeologous loci, the resulting PCR product contained a mixture of sequences from both the A- and the D-subgenomes. To separate these products into their respective subgenomic sequences, we recovered the PCR products with Geneclean II (Bio 101), ligated the PCR products into pGEM-

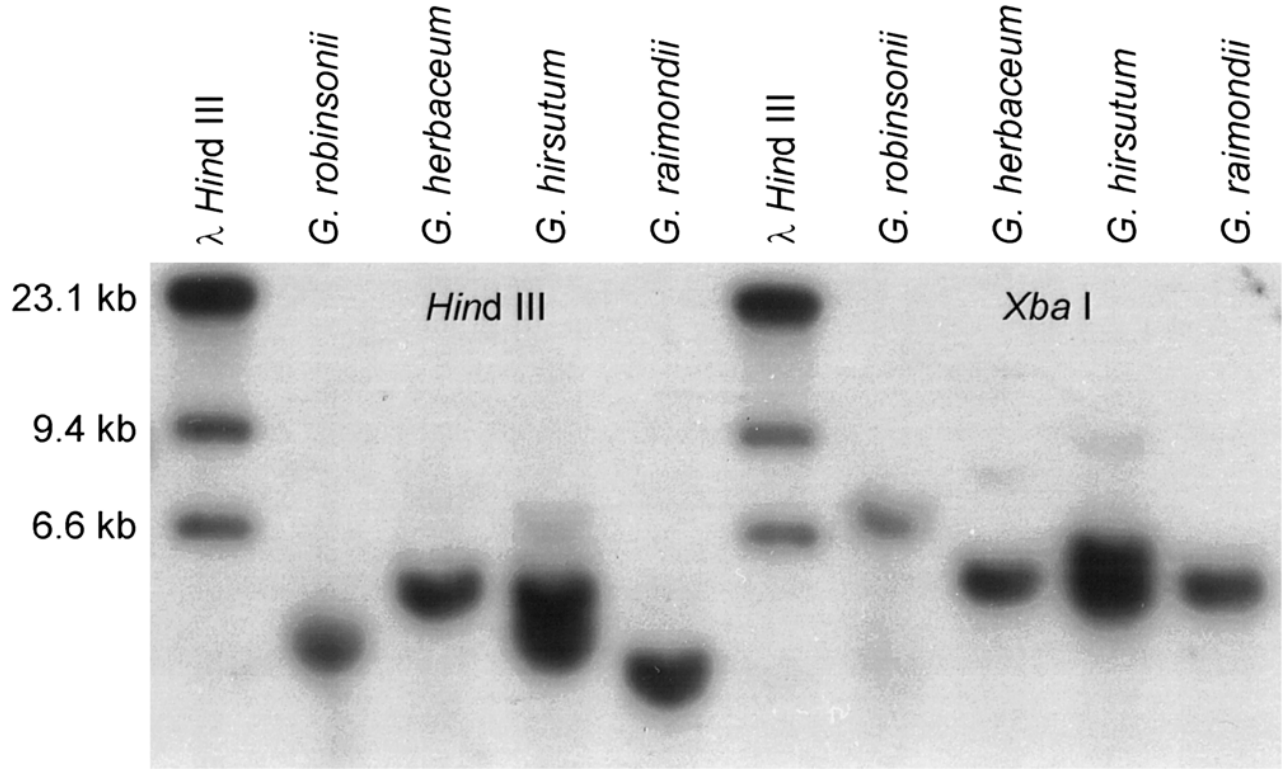

FIG. 2.—Southern blot of an *AdhA*-specific probe hybridized to *Hin*dIII and *Xba*I digested genomic DNAs of three diploid (*G. robinsonii*, C-genome; *G. herbaceum*, A-genome; *G. raimondii*, D-genome) and one tetraploid (*G. hirsutum*) cotton species. In both sets of digests, the probe reveals only a single band per diploid genome, indicating that *AdhA* is single-copy.

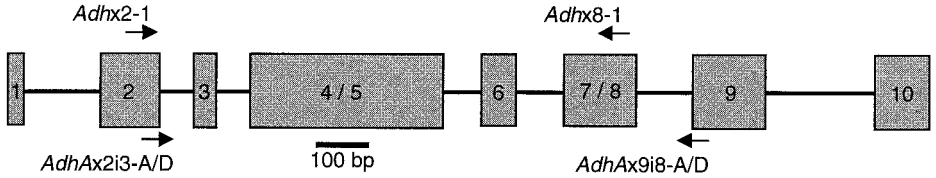

FIG. 3.—Diagrammatic representation of the *Gossypium AdhA* locus. Exons are shown as shaded boxes, introns as the lines connecting the exons. Genomic sequence data are available only for exons 2–8; the lengths of exons 1, 9, and 10 and introns 1 and 9 are extrapolated from other *Gossypium Adh* sequences. PCR amplification primers are shown in their approximate positions. *AdhA*-specific primer pair *Adh*x2-1 and *Adh*x8-1 is shown above the diagram, while homoeolog-specific primer pairs (*AdhA*x2i3-A + *AdhA*x8i9-A; *AdhA*x2i3-D + *AdhA*x9i8-D) are shown below the diagram. A 100-bp scale bar is included for reference.

T (Promega), and transformed competent *Escherichia coli* Top10 F′ cells (Invitrogen). Resulting colonies were screened for inserts by resuspending bacterial colonies in 10 μl of water, boiling for 10 min, centrifuging for 30 s, and using 2.5 μl of the supernatant as a template in a 10-μl PCR reaction using the original amplification primers and reaction conditions. PCR products that were the correct size (indicating presence of an *AdhA* insert) were ethanol-precipitated, resuspended in 10 μl of water, and subjected to restriction digestion with *Taq*$^{\alpha}$I (New England Biolabs), an enzyme that has one recognition site in sequences from the A-subgenome and two sites in sequences from the D-subgenome. This procedure allowed us to distinguish the subgenomic origin of each individual clone. To eliminate sequencing artifacts caused by misincorporation during PCR, for each accession, we isolated and pooled 10 plasmids from each subgenome and sequenced this pool using the amplification primers as sequencing primers. For 39 of 48 templates (81%), this procedure resulted in a monomorphic sequencing ladder; i.e., no apparent heterozygotes were detected. The remaining templates showed polymorphism at one or more sites. To evaluate whether these polymorphisms reflected true heterozygosity or misincorporation, we repeated the amplification, cloning, and sequencing steps. For all sequences that were initially based on clones, we were unable to reproduce the polymorphisms detected earlier, and in several cases, new polymorphisms appeared. We concluded that in all but one case, polymorphisms did not reflect true heterozygosity but instead arose due to PCR error.

After the initial results were obtained, a second strategy was employed to isolate subgenome-specific *AdhA* sequences. This approach involved the use of two pairs of homoeolog-specific PCR primers that would amplify *AdhA* from only one subgenome at a time (A-subgenome-specific primers—*AdhA*x2i3-A: AAG GTA TTA CTG TAC GAT AA; *AdhA*x9i8-A: CCT GTA ATT CAA GAA GAA G; D-subgenome-specific primers—*AdhA*x2i3-D: AAG GTA TTA CTG TTC GAT AT; *AdhA*x9i8-D: CCT GTA ATT CAA GAA GCA T). These primers generated homogeneous PCR pools that could be directly sequenced, obviating the laborious cloning and restriction digestion steps. In addition, direct sequencing of PCR products (as opposed to sequencing of cloned PCR products) greatly reduces the likelihood of detecting misincorporated nucleotides, since these are expected to be present in low concentrations relative to the correct products. Therefore, we reamplified *AdhA* (using the homoeolog-specific primers) and sequenced the PCR products directly from those accessions that had shown polymorphism. Sequences obtained using this approach were monomorphic, indicating that we had eliminated PCR artifacts.

All DNA sequencing was performed using the Thermo Sequenase ^{33}P-radiolabeled terminator cycle sequencing kit (Amersham). Sequencing reactions were electrophoresed on 5%–6% Long Ranger sequencing gels (FMC). Because so little polymorphism was detected, templates were sequenced on one strand only. After the entire data set had been collected, each sequence was rechecked at all polymorphic sites to confirm the original reads. The sequences reported here have been submitted to GenBank under accession numbers AF085064–AF085085, AF085812–AF085821, and AF090146–AF090168.

Statistical Analyses

Gossypium hirsutum and *G. barbadense* are genomic allotetraploids and display disomic inheritance. For analytical purposes, we assumed that our approach detected both alleles, and we therefore represented each locus in each accession by the two alleles present notwithstanding the high level of homozygosity observed (cf. Gaut and Clegg 1993*b*). Our experimental design of pooling 10 plasmids per homoeolog for each accession was designed to eliminate *Taq* error, as well as to ensure cloning and sequencing of both alleles. Assuming equal representation of both alleles in the PCR product pool and equivalent success of cloning each allele, the probability of not including both alleles in the plasmid pool is quite small (0.5^{10} or $P = 0.001$). In our experience, even if an allele is represented only once in the plasmid pool, it would be detected in the sequencing ladder. Finally, in other studies of *AdhA* in diploid *Gossypium* species, we used identical PCR primers and readily amplified both alleles of heterozygous individuals. The foregoing observations suggest that our approach is expected to detect both alleles at a locus and that the monomorphic sequencing ladders we obtained were the result of homozygosity.

For each subgenome of both *G. hirsutum* and *G. barbadense*, we calculated two measures of nucleotide diversity per base pair: π (Nei 1987, pp. 256–257) and θ_w (Watterson 1975). The former measure quantifies the mean percentage of nucleotide differences among all pairwise comparisons for a set of sequences, whereas the latter is simply an index of the number of segregat-

ing (polymorphic) sites. Under neutral expectations, θ_w is equal to π (Tajima 1989, 1993*b*). A 95% confidence interval around θ_w was calculated for *AdhA* from both subgenomes of *G. hirsutum* using methods described by Kreitman and Hudson (1991). Tests of neutral evolution were performed as described by Tajima (1989), Fu and Li (1993), and Hudson, Kreitman, and Aguadé (1987). Recombination was explored using the algorithm of Hudson and Kaplan (1985). Many of the above calculations were expedited by the software program DnaSP, version 2.52 (Rozas and Rozas 1997). Estimates of genetic diversity (mean number of alleles per locus—*A*; mean panmictic [expected] heterozygosity—$H_T = 1 - \Sigma[p_i]^2$, where p_i represents allele frequencies; cf. Brubaker and Wendel 1994) were calculated using our sequence data as well as previously published isozyme (Wendel, Brubaker, and Percival 1992) and RFLP (Brubaker and Wendel 1994) data for a comparable set of accessions (identical accessions for RFLP data; three missing accessions for isozyme data).

Given the phylogenetic framework of the genus *Gossypium* and estimates of the timing of several major branching points in the phylogeny (fig. 1; Wendel and Albert 1992; Seelanan, Schnabel, and Wendel 1997), we were able to estimate an absolute mutation rate for *AdhA*. Specifically, using unpublished *AdhA* sequences of *G. robinsonii* (C-genome outgroup), *G. herbaceum* (A-genome diploid), and *G. raimondii* (D-genome diploid), we generated, using exon data only, a synonymous-site Jukes-Cantor (JC) distance matrix using MEGA, version 1.0 (Kumar, Tamura, and Nei 1993). The timings of the two branchpoint estimates shown in figure 1 were derived from analyses of chloroplast *ndhF* sequences (Seelanan, Schnabel, and Wendel 1997). Using these divergence time points and the JC distances, we estimated the absolute synonymous mutation rate as the JC distance divided by twice the time since divergence.

Results

DNA Polymorphism

We determined approximately 1 kb of sequence from both the A- and the D-subgenomic homoeologs of *AdhA* for 22 accessions (44 alleles per subgenome) of *G. hirsutum* and 5 accessions (10 alleles per subgenome) of *G. barbadense*. Thus, approximately 108 kb of effective sequence data were generated (27 accessions $\times$ 2 alleles/homoeolog $\times$ 2 homoeologs). Sequence data for each allele consist of 662 bp of coding sequence and 336 bp of intron sequence; this represents a mean of 500.33 nonsynonymous sites and 482.67 silent sites (synonymous or intron, excluding gaps).

All sequences appeared homozygous, with the exception of one *G. hirsutum* cultivar (Paymaster H86048) which was heterozygous for alleles 1D and 2D (see fig. 4). The distribution of nucleotides at all polymorphic sites for both homoeologs is shown in figure 4. In the A-subgenome of *G. hirsutum*, we observed only one polymorphic site (position 571), which included approximately equal representation among accessions of the alternative nucleotides G and A. This transitional and silent substitution was at a third codon position. In the A-subgenome of *G. barbadense*, no polymorphic sites were observed. In the D-subgenome of *G. hirsutum*, there were three polymorphic sites, all within introns. Two of these three sites (positions 84 and 942) reflected transitional mutations, while the third polymorphism resulted from a [G, T] transversion (position 684). For all three polymorphic sites, the minority state occurred in either 5 or 6 of the 22 accessions sampled. One polymorphic site was revealed in the D-subgenome of *G. barbadense* (position 511, a third codon position transition).

No nucleotides at either *AdhA* homoeolog distinguish all *G. barbadense* alleles from those of *G. hirsutum*. For the A-subgenome locus, all five *G. barbadense* accessions are homozygous for an allele shared by 8 of the 22 *G. hirsutum* accessions. Similarly, for *AdhA* from the D-subgenome, four of the five *G. barbadense* accessions are homozygous for an allele shared by 11 of the 22 *G. hirsutum* accessions (fig. 4).

Estimates of nucleotide diversity (π, θ_w; gaps treated as missing data) are shown for each data set in table 2. These estimates show that nucleotide diversity is approximately twice as high for *AdhA* from the D-subgenome as it is for *AdhA* from the A-subgenome.

The two homoeologs of *AdhA* differed by a minimum of 20 nucleotide substitutions within both *G. hirsutum* and *G. barbadense*, representing 12 transitions and 8 transversions (fig. 4). Thus, the two *AdhA* homoeologs exhibit approximately 2% sequence divergence based on nongapped positions. In addition, the two homoeologs are differentiated by four gaps, all of which are intron nucleotides present in the A-subgenomic homoeologs that are absent from the corresponding locus in the D-subgenome. All available data indicate that the *AdhA* homoeologs have evolved independently subsequent to polyploid formation; i.e., there is no evidence of intersubgenomic gene conversion or recombination. This inference is supported by the 20 nucleotide substitutions and 4 gaps that distinguish the homoeologs, the majority of which are also shared with the respective diploid progenitors.

Tests of Neutral Evolution, Recombination, and Rates of Nucleotide Substitution

Several statistical tests were used to test the hypothesis that *AdhA* sequences have been evolving in accordance with expectations under neutral theory. The tests of Tajima (1989) and Fu and Li (1993) compare different estimates of θ ($4N_e\mu$), each of which makes certain assumptions about how sequences evolve (Simonsen, Churchill, and Aquadro 1995; Wayne and Simonsen 1998). These tests were conducted on each of the four data sets (two subgenomes in two species), and the results are shown in table 2. None of these tests returned significant *P* values. This is not surprising, given the small number of variable positions and the relatively low statistical power of these tests (Simonsen, Churchill, and Aquadro 1995; Wayne and Simonsen 1998). The HKA test (Hudson, Kreitman, and Aguadé

	Accession	35	41	77	84	89	165	180	184	189	202	222	271	283	454	511	514	571	580	662	684	804	823	824	825	826	827	828	829	830	845	859	867	868	942	945	948	Allele
A-subgenome	pfx-A	A	A	T	G	T	A	T	A	T	A	G	G	T	T	C	G	A	C	T	G	A	C	A	T	T	A	T	A	T	C	T	C	T	C	A	T	1A
G. hirsutum	TX-1-A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	A	T	T	A	T	A	T	.	.	.	.	.	.	.	1A
	TX-6-A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	A	T	T	A	T	A	T	.	.	.	.	.	.	.	1A
	TX-21-A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	A	T	T	A	T	A	T	.	.	.	.	.	.	.	1A
	TX-44-A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	A	T	T	A	T	A	T	.	.	.	.	.	.	.	1A
	TX-51-A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	A	T	T	A	T	A	T	.	.	.	.	.	.	.	1A
	TX-93-A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	A	T	T	A	T	A	T	.	.	.	.	.	.	.	1A
	TX-94-A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	C	A	T	T	A	T	A	T	.	.	.	.	.	.	.	2A
	TX-98-A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	A	T	T	A	T	A	T	.	.	.	.	.	.	.	1A
	TX-111-A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	C	A	T	T	A	T	A	T	.	.	.	.	.	.	.	2A
	TX-116-A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	A	T	T	A	T	A	T	.	.	.	.	.	.	.	1A
	TX-119-A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	C	A	T	T	A	T	A	T	.	.	.	.	.	.	.	2A
	TX-166-A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	C	A	T	T	A	T	A	T	.	.	.	.	.	.	.	2A
	TX-188-A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	A	T	T	A	T	A	T	.	.	.	.	.	.	.	1A
	TX-192-A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	A	T	T	A	T	A	T	.	.	.	.	.	.	.	1A
	TX-367-A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	C	A	T	T	A	T	A	T	.	.	.	.	.	.	.	2A
	TX-481-A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	C	A	T	T	A	T	A	T	.	.	.	.	.	.	.	2A
	TX-706-A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	C	A	T	T	A	T	A	T	.	.	.	.	.	.	.	2A
	TX-766-A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	C	A	T	T	A	T	A	T	.	.	.	.	.	.	.	2A
	Paymaster-A	?	?	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	A	T	T	A	T	A	T	.	.	.	.	.	.	.	1A
	Deltapine-A	?	?	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	A	T	T	A	T	A	T	.	.	.	.	.	.	.	1A
	BR115-A	?	?	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	A	T	T	A	T	A	T	.	.	.	.	.	.	.	1A
G. barbadense	PimaS5-A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	C	A	T	T	A	T	A	T	.	.	.	.	.	.	.	2A
	B106-A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	C	A	T	T	A	T	A	T	.	.	.	.	.	.	.	2A
	B250-A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	C	A	T	T	A	T	A	T	.	.	.	.	.	.	.	2A
	B444-A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	C	A	T	T	A	T	A	T	.	.	.	.	.	.	.	2A
	B559-A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	C	A	T	T	A	T	A	T	.	.	.	.	.	.	.	2A
D-subgenome	pfx-D	T	T	-	A	C	-	-	G	C	T	A	A	C	C	.	T	G	T	A	.	C	-	-	-	-	-	-	-	-	T	C	T	G	.	G	A	1D
G. hirsutum	TX-1-D	T	T	-	A	C	-	-	G	C	T	A	A	C	C	.	T	G	T	A	.	C	-	-	-	-	-	-	-	-	T	C	T	G	.	G	A	1D
	TX-6-D	T	T	-	.	C	-	-	G	C	T	A	A	C	C	.	T	G	T	A	.	C	-	-	-	-	-	-	-	-	T	C	T	G	.	G	A	2D
	TX-21-D	T	T	-	.	C	-	-	G	C	T	A	A	C	C	.	T	G	T	A	T	C	-	-	-	-	-	-	-	-	T	C	T	G	T	G	A	3D
	TX-44-D	T	T	-	.	C	-	-	G	C	T	A	A	C	C	.	T	G	T	A	.	C	-	-	-	-	-	-	-	-	T	C	T	G	.	G	A	2D
	TX-51-D	T	T	-	.	C	-	-	G	C	T	A	A	C	C	.	T	G	T	A	T	C	-	-	-	-	-	-	-	-	T	C	T	G	T	G	A	3D
	TX-93-D	T	T	-	.	C	-	-	G	C	T	A	A	C	C	.	T	G	T	A	.	C	-	-	-	-	-	-	-	-	T	C	T	G	.	G	A	2D
	TX-94-D	T	T	-	.	C	-	-	G	C	T	A	A	C	C	.	T	G	T	A	T	C	-	-	-	-	-	-	-	-	T	C	T	G	T	G	A	3D
	TX-98-D	T	T	-	.	C	-	-	G	C	T	A	A	C	C	.	T	G	T	A	.	C	-	-	-	-	-	-	-	-	T	C	T	G	.	G	A	2D
	TX-111-D	T	T	-	.	C	-	-	G	C	T	A	A	C	C	.	T	G	T	A	T	C	-	-	-	-	-	-	-	-	T	C	T	G	T	G	A	3D
	TX-116-D	T	T	-	.	C	-	-	G	C	T	A	A	C	C	.	T	G	T	A	.	C	-	-	-	-	-	-	-	-	T	C	T	G	.	G	A	2D
	TX-119-D	T	T	-	A	C	-	-	G	C	T	A	A	C	C	.	T	G	T	A	.	C	-	-	-	-	-	-	-	-	T	C	T	G	.	G	A	1D
	TX-166-D	T	T	-	.	C	-	-	G	C	T	A	A	C	C	.	T	G	T	A	.	C	-	-	-	-	-	-	-	-	T	C	T	G	.	G	A	2D
	TX-188-D	T	T	-	.	C	-	-	G	C	T	A	A	C	C	.	T	G	T	A	.	C	-	-	-	-	-	-	-	-	T	C	T	G	.	G	A	2D
	TX-192-D	T	T	-	.	C	-	-	G	C	T	A	A	C	C	.	T	G	T	A	.	C	-	-	-	-	-	-	-	-	T	C	T	G	.	G	A	2D
	TX-367-D	T	T	-	A	C	-	-	G	C	T	A	A	C	C	.	T	G	T	A	.	C	-	-	-	-	-	-	-	-	T	C	T	G	.	G	A	1D
	TX-481-D	T	T	-	A	C	-	-	G	C	T	A	A	C	C	.	T	G	T	A	.	C	-	-	-	-	-	-	-	-	T	C	T	G	.	G	A	1D
	TX-706-D	T	T	-	.	C	-	-	G	C	T	A	A	C	C	.	T	G	T	A	.	C	-	-	-	-	-	-	-	-	T	C	T	G	T	G	A	4D
	TX-766-D	T	T	-	.	C	-	-	G	C	T	A	A	C	C	.	T	G	T	A	T	C	-	-	-	-	-	-	-	-	T	C	T	G	T	G	A	3D
	Paymaster-D	?	?	-	R	C	-	-	G	C	T	A	A	C	C	.	T	G	T	A	.	C	-	-	-	-	-	-	-	-	T	C	T	G	.	G	A	1D/2D
	Deltapine-D	?	?	-	.	C	-	-	G	C	T	A	A	C	C	.	T	G	T	A	.	C	-	-	-	-	-	-	-	-	T	C	T	G	.	G	A	2D
	BR115-D	?	?	-	.	C	-	-	G	C	T	A	A	C	C	.	T	G	T	A	.	C	-	-	-	-	-	-	-	-	T	C	T	G	.	G	A	2D
G. barbadense	PimaS5-D	T	T	-	.	C	-	-	G	C	T	A	A	C	C	.	T	G	T	A	.	C	-	-	-	-	-	-	-	-	T	C	T	G	.	G	A	2D
	B106-D	T	T	-	.	C	-	-	G	C	T	A	A	C	C	.	T	G	T	A	.	C	-	-	-	-	-	-	-	-	T	C	T	G	.	G	A	2D
	B250-D	T	T	-	.	C	-	-	G	C	T	A	A	C	C	T	T	G	T	A	.	C	-	-	-	-	-	-	-	-	T	C	T	G	.	G	A	5D
	B444-D	T	T	-	.	C	-	-	G	C	T	A	A	C	C	.	T	G	T	A	.	C	-	-	-	-	-	-	-	-	T	C	T	G	.	G	A	2D
	B559-D	T	T	-	.	C	-	-	G	C	T	A	A	C	C	.	T	G	T	A	.	C	-	-	-	-	-	-	-	-	T	C	T	G	.	G	A	2D

FIG. 4.—Polymorphic nucleotide positions in the *AdhA* data set. At each polymorphic nucleotide site (numbers shown above the sequences), the nucleotide state observed in each accession is given relative to the *G. hirsutum* pfx-A sequence. A period denotes identity, while question marks denote missing data. Alignment gaps are indicated by dashes. Allelic designations are given in the final column.

1987) compares levels of polymorphism between genes or regions both within and between species, the assumption being that levels of neutral polymorphism should be correlated with rates of evolution across genomes. While the original intent of this test was to compare an unknown region with a region that is presumed to be evolving neutrally, we adapted it to test the assumption that the two homoeologs are evolving equivalently. Intraspecific polymorphism at the *G. hirsutum* *AdhA* A-subgenome homoeolog was compared with that

Table 2
Estimates ($\times 10^3$) of Nucleotide Diversity per Base Pair (π, θ_w) and Tests of Neutral Evolution

	π	θ_w (θ_L, θ_U)[a]	D[b]	D[c]	F[c]
Gossypium hirsutum A-subgenome	0.50	0.24 (0.0006, 1.52)	1.47	0.55	0.94
G. hirsutum D-subgenome	1.23	0.74 (0.16, 2.38)	1.42	0.91	1.24
Gossypium barbadense A-subgenome	0.00	0.00	—	—	—
G. barbadense D-subgenome	0.36	0.36 (0.009, 2.73)	0.62	0.74	0.67

[a] Lower (θ_L) and upper (θ_U) bounds ($\times 10^3$) of the 95% confidence intervals in parentheses.
[b] Test statistic of Tajima (1989); no results are statistically significant.
[c] Test statistics of Fu and Li (1993); no results are statistically significant.

at the *AdhA* D-subgenome homoeolog; the same regions from *G. barbadense* provided the interspecific comparison. The HKA test result was not significant (P = 0.75). The Hudson and Kaplan (1985) estimate of the minimum number of recombination events was zero, viz., all sites were compatible with a history devoid of interallelic recombination. A network depicting allele relationships and their corresponding taxonomic and geographic distribution is shown in figure 5.

An absolute rate of nucleotide substitution was estimated for *AdhA* using two separate calibration points (fig. 1) derived from analyses of chloroplast DNA sequence data (Seelanan, Schnabel, and Wendel 1997). Using the divergence of the [A+D]-genome clade from the C-genome clade (synonymous-site JC distance = 0.035; divergence time of 12 Myr), a rate of 1.47×10^{-9} synonymous substitutions per synonymous site per year was obtained. Using the split between the A- and D-genomes (synonymous-site JC distance = 0.045; divergence time of 11 Myr), the substitution rate was estimated to be 2.05×10^{-9} synonymous substitutions per synonymous site per year.

Comparisons of Measures of Genetic Diversity

One difference between DNA sequencing and indirect methods of assessing genetic variation (e.g., isozymes or RFLPs) is that all polymorphisms will be detected in the region sequenced, not just those that result in restriction site mutations (cf. RFLP analysis) or electrophoretically detectable charge or conformational changes (cf. isozyme analysis). Thus, one might expect levels of allelic diversity and heterozygosity to be higher for nucleotide sequence data than for other data sets; this expectation was met in the present study (table 3). Previous studies have assayed isozyme and RFLP diversity in *G. hirsutum* and *G. barbadense* (Wendel, Brubaker, and Percival 1992; Brubaker and Wendel 1994). We recalculated genetic diversity statistics for the isozyme and RFLP data by pruning the data sets to include only those accessions sampled here (table 3). In general, allelic diversity was higher for *AdhA* sequence data than for isozymes or RFLPs, as expected. In addition, expected heterozygosity (=mean panmictic heterozygosity) was also higher for the sequence data, but observed heterozygosity at *AdhA* was zero in all cases except for the D-subgenome of *G. hirsutum.* The single heterozygote observed was for a cultivar (Paymaster H86048); heterozygosity in this accession may reflect the results of a breeding program or germplasm maintenance.

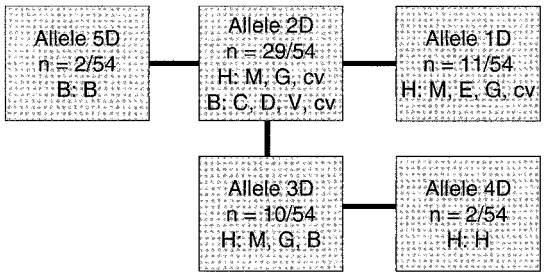

FIG. 5.—Allele network depicting relationships among alleles observed at the D-subgenome homoeolog of *G. hirsutum* and *G. barbadense*; each allele differs from alleles to which it is connected by a single nucleotide substitution. Allele designations follow figure 4. For each allele, the number of times it was observed out of a total of 54 alleles (44 *G. hirsutum* alleles and 10 *G. barbadense* alleles) is given. The taxonomic and geographic distributions of alleles are as follows: H—allele detected in *G. hirsutum*; B—allele detected in *G. barbadense*; B—Belize; C—Colombia; D—Dominican Republic; E—El Salvador; G—Guatemala; H—Honduras; M—Mexico; V—Venezuela; cv—cultivar.

Discussion

Nucleotide Diversity in Allopolyploid *Gossypium*

A primary conclusion of the present study is that nucleotide diversity for *AdhA* in *G. hirsutum* and *G. barbadense* is very low. Estimates reported here are lower

Table 3
Genetic Diversity Statistics for Isozyme, RFLP, and *AdhA* Sequence Data

	A[a]	H_T[b]	Obs. Het.[c]
Gossypium hirsutum			
Isozymes	1.4	0.126	0.006
RFLPs	1.6	0.144	0.004
AdhA, A-subgenome	2	0.463	0.000
AdhA, D-subgenome	4	0.656	0.045
AdhA loci, mean	3	0.556	0.023
Gossypium barbadense			
Isozymes	1.2	0.074	0.000
RFLPs	1.2	0.062	0.008
AdhA, A-subgenome	1	0.000	0.000
AdhA, D-subgenome	2	0.320	0.000
AdhA loci, mean	1.5	0.160	0.000

[a] Mean number of alleles per locus.
[b] Mean panmictic (expected) heterozygosity.
[c] Observed heterozygosity (number of heterozygous accessions/total number of accessions sampled).

than previously reported values not only for plant *Adh* sequences (see table 3 of Cummings and Clegg 1998; Liu, Zhang, and Charlesworth 1998), but for other plant nuclear genes as well (*C1* in maize—Hanson et al. 1996; *ChiA* in *Arabidopsis*—Kawabe et al. 1997; *ChsA* in *Ipomoea*—Huttley et al. 1997; *Pgi* in *Dioscorea*—Terauchi, Terachi, and Miyashita 1997). Nucleotide diversity per base pair for nuclear genes in other plant species ranges from a low of $\theta_w = 0.001$ at *Pgi* in *Dioscorea* (Terauchi, Terachi, and Miyashita 1997) to a high of $\theta_w = 0.025$ at *Adh1* in *Zea mays* ssp. *parviglumus* (Eyre-Walker et al. 1998); our values of θ_w ranged from 0.000 (*G. barbadense* A-subgenome) to 0.0007 (*G. hirsutum* D-subgenome).

Potential explanations for such low levels of nucleotide diversity include one or more recent genetic bottlenecks, a low mutation rate, a self-pollinating reproductive biology, and a selective sweep. We present evidence that the first three of these factors have been important in shaping the population genetic structure of cotton and are sufficient to explain our observations, thus obviating the need to invoke additional mechanisms such as selective sweeps. Operating together, the historical and life history features of *G. hirsutum* have had a net effect of severely constraining levels of genetic diversity, as discussed in the following paragraphs.

Genetic Bottlenecks

Accumulated evidence indicates that *G. hirsutum* and *G. barbadense* are allotetraploids that are derived from a single polyploidization event that occurred ~1–2 MYA (Wendel 1989; Seelanan, Schnabel, and Wendel 1997; Small et al. 1998). Because the two parental diploid genomes are confined to different continents (A-genome: Africa-Asia; D-genome: New World, primarily Mexico), polyploidization appears to have been precipitated by transoceanic dispersal of an A-genome propagule to the New World, followed by hybridization and allopolyploidization with the native D-genome donor. It seems likely that only one to a few A-genome propagules made this transoceanic voyage, and it seems similarly probable that only one to very few individuals were involved in the initial hybridization event from which allopolyploid *Gossypium* emerged. Thus, the process by which the lineage formed is characterized by a severe genetic bottleneck; presumably, one or a few hybrid individual(s) constituted the entire gene pool from which the extant tetraploids are derived. Subsequent diversification of the nascent allopolyploid into the five modern tetraploid species implicates additional genetic bottlenecks associated with these more recent speciation events. Finally, more recent bottlenecks undoubtedly occurred as a consequence of the domestication of *G. hirsutum*, perhaps 4,000–5,000 years ago (Wendel, Brubaker, and Percival 1992; Brubaker and Wendel 1994; Wendel 1995).

Thus, the agronomic development of modern *G. hirsutum* varieties has been characterized by sequential genetic bottlenecks followed by rapid range expansions. A similar history with an approximately equivalent timescale has been described for *G. barbadense* (Percy and Wendel 1990). For both species, these episodic bottlenecks undoubtedly contributed to a winnowing of nucleotide diversity, which may not have been especially extensive even in the wild progenitors. This winnowing process has occurred over a period of time that is exceptionally brief on an evolutionary timescale, especially in light of the time frame necessary for the introduction of genetic diversity through mutation.

Breeding System

An additional constraint on nucleotide diversity levels in *G. hirsutum* and *G. barbadense* stems from their reproductive biology; both species are self-compatible and produce a high proportion of their seed through self-pollination (Wendel 1995). Self-pollination is known to reduce effective population size, which in turn reduces expected levels of genetic diversity (Pollak 1987; Liu, Zhang, and Charlesworth 1998). In addition to reducing levels of genetic diversity, selfing is expected to reduce observed heterozygosity relative to expected heterozygosity, as allelic variation manifests itself as alternative homozygotes rather than heterozygotes. Thus, the breeding system is consistent with our observation of a near-complete absence of observed heterozygosity at *AdhA*, where the sole exception was for a *G. hirsutum* cultivar that may have acquired its heterozygosity (either intentionally or unintentionally) through a breeding program or during seed increase for germplasm maintenance.

Low Mutation Rate

The absolute synonymous substitution rate calculated for *AdhA* in *Gossypium* is 1.47×10^{-9} to 2.05×10^{-9} substitutions per site per year. Wolfe, Li, and Sharp (1987) estimated that synonymous substitution rates at plant nuclear genes range from 5×10^{-9} to 30×10^{-9} and average 5.1×10^{-9} to 7.1×10^{-9} (Wolfe, Sharp, and Li 1989). Gaut (1998) has estimated a synonymous rate of 6.03×10^{-9} in a comparison of nine nuclear genes of rice and maize. The lowest published synonymous rate for a plant nuclear gene is 2.61×10^{-9} for *AdhA* in palms (Gaut et al. 1996). The synonymous substitution rate for *AdhA* in *Gossypium* is therefore 2.5 to 4 times lower average rates and is lower than any previously published rates. This estimate is, in fact, within the range (1.0×10^{-9} to 3.0×10^{-9}) of synonymous substitution rates in chloroplast genes (Wolfe, Li, and Sharp 1987). Given this slow mutation rate, there has been little time for the genesis of allelic diversity since species formation (perhaps 1–2 Myr), and even less time since *G. hirsutum* and *G. barbadense* were domesticated (perhaps 4,000–5,000 years). Even under a scenario of complete retention of genetic diversity, i.e., no loss due to sampling or genetic bottlenecks (as discussed above), the expectation is that for *AdhA*, with approximately 500 silent sites and mutation rates as estimated above, only one or two nucleotides, on average, are expected to become polymorphic in each million years. Hence, the observation of only one and three polymorphic sites in the A- and D-subgenomic homoeologs, respectively, is consistent with expectations based on our understanding of

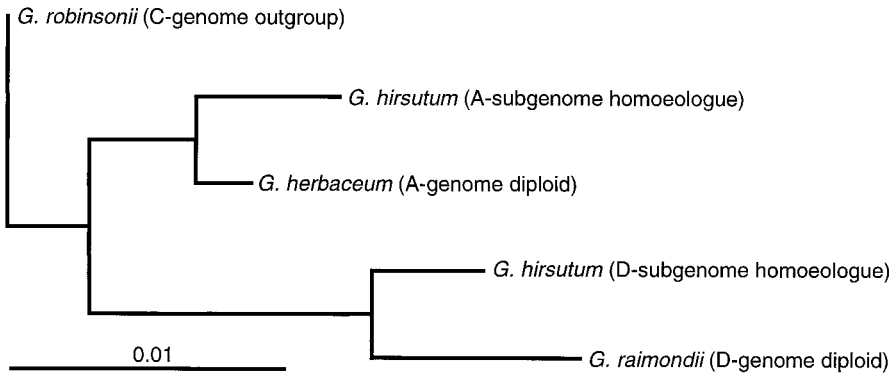

FIG. 6.—Neighbor-joining tree of *AdhA* sequences from diploid and tetraploid *Gossypium,* rooted with *G. robinsonii* (the topology resulting from maximum-parsimony analysis is identical). Homoeologs from allotetraploid *G. hirsutum* (accession pfx) cluster with their respective diploid progenitors. The branch leading to the *G. hirsutum* D-homoeolog and *G. raimondii* (D-diploid) is ca. twice as long as the branch leading to the *G. hirsutum* A-homoeolog and *G. herbaceum* (A-diploid). Tajima's (1993*a*) 1D relative-rate test returns a statistically significant rate difference only in the comparison of *G. raimondii* and *G. herbaceum.*

mutation rates and the history and biology of the species. It therefore seems unnecessary to invoke additional mechanisms such as selective sweeps.

Lack of Coalescence

One of the noteworthy observations of this study is that *AdhA* alleles do not coalesce within species. In both the A- and the D-subgenomes, the predominant allele found in *G. barbadense* also occurs at high frequency in *G. hirsutum.* This result is consistent with the low mutation rates and phylogenetic history discussed above or, alternatively, with a hypothesis of large-scale introgression of *G. hirsutum* alleles into *G. barbadense* (Brubaker, Koontz, and Wendel 1993). With respect to the former, molecular phylogenetic data have led to the suggestion that postpolyploidization, there was a rapid radiation into the extant clades represented by the five modern species (fig. 1; Wendel 1989; Small et al. 1998). Under such circumstances (low variation and rapid radiation), it is expected that alleles would be shared across species boundaries, unless a high mutation rate and a high fixation rate were operating.

Comparison Within and Between Homoeologous Locus Pairs

One of the initial goals of this study was to test the hypothesis that homoeologous loci exhibit equivalent evolutionary dynamics. Given a single origin of the tetraploid *Gossypium* species, levels and patterns of genetic diversity should be equivalent for homoeologous loci, assuming an absence of selection, differential recombination, or other forces that might differentially affect members of a homoeologous locus pair. All population-level factors other than selection (e.g., effective population size, genetic drift, breeding system) are equivalent.

A previous study (Small et al. 1998) has shown that for another alcohol dehydrogenase locus in *Gossypium* (*AdhC*), evolutionary rates at the two homoeologs differ significantly, with the locus from the D-genome diploid and D-subgenome of the tetraploids evolving at a faster rate. Neutral theory predicts that evolutionary rate and genetic diversity should be positively correlated—this is, in fact, the basis of the HKA test (Hudson, Kreitman, and Aguadé 1987). We applied this test to the *AdhA* data presented here, not to detect departure from neutrality, but to detect differences in evolutionary dynamics between homoeologs; the result was not significant. Likewise, the 95% confidence intervals calculated for θ_w largely overlap. Finally, application of Tajima's (1993*a*) 1D relative-rate test comparing *AdhA* sequences from the A-genome diploid and the A-subgenome of *G. hirsutum* with those from the D-diploid and D-subgenome of *G. hirsutum* returned only one significant departure from rate homogeneity (*G. herbaceum* vs. *G. raimondii*), despite a qualitatively obvious rate difference (fig. 6). Although none of the statistical tests supports an inference of rate inequality among the *AdhA* homoeologs, allelic diversity is twice as high in the D-subgenome, and nucleotide diversity is two to three times higher in the D-subgenome, results that are directionally consistent with the previously reported *AdhC* data (Small et al. 1998). These observations may or may not be consequential; data from other homoeologous pairs are needed to evaluate the possibility that the subgenomes of *G. hirsutum* are subject to different evolutionary pressures.

Comparisons among homoeologous locus pairs may also provide insight into processes of genomic evolution. The evolutionary dynamics appear to be different for *AdhA* and *AdhC* in tetraploid *Gossypium.* For example, in the tetraploid species of *Gossypium,* sequence divergence for *AdhA* (A-subgenome vs. D-subgenome) averages 2%, but it is over 4% for *AdhC* (Small et al. 1998). As described above, increased evolutionary rate should be correlated with increased genetic diversity, which would predict a higher level of diversity at *AdhC* than at *AdhA.* We are currently conducting studies to test this hypothesis. Finally, previous studies have shown that a correlation exists between chromosomal position (and associated rates of recombination) and levels of genetic diversity at a locus. In general, the more distal a locus is from the centromere, the higher the

recombination rate and genetic diversity will be (Begun and Aquadro 1992; Aquadro and Begun 1993; Dvorák, Luo, and Yang 1998). Genetic mapping data allow us to speculate that *Gossypium Adh* loci may show the opposite trend. The slowly evolving, low-diversity locus *AdhA* resides at the distal end of a linkage group, while the quickly evolving locus *AdhC* maps near the middle of a linkage group (unpublished data). While we are currently unable to correlate these genetic mapping data with a physical map and therefore pinpoint distances from the centromere or telomere, these preliminary data may provide an exception to the general relationship between genetic diversity and chromosomal position.

Acknowledgments

We thank C. Brubaker for his previous work on genetic diversity and mapping in tetraploid *Gossypium*, which greatly facilitated this investigation. We also thank B. Gaut for helpful suggestions and discussion. We thank C. Brubaker, R. Cronn, J. Doebley, B. Gaut, and P. Soltis for comments on an earlier draft of this manuscript. This research was supported by funding from the National Science Foundation (to J.F.W.).

LITERATURE CITED

Amos, W., and J. Harwood. 1998. Factors affecting levels of genetic diversity in natural populations. Philos. Trans. R. Soc. Lond. B Biol. Sci. **353**:177–186.

Aquadro, C. F., and D. J. Begun. 1993. Evidence for and implications of genetic hitchhiking in the *Drosophila* genome. Pp. 159–178 *in* N. Takahata and A. G. Clark, eds. Mechanisms of molecular evolution. Sinauer, Sunderland, Mass.

Begun, D. J., and C. F. Aquadro. 1992. Levels of naturally occurring DNA polymorphism correlate with recombination rates in *D. melanogaster*. Nature **356**:519–520.

Bergelson, J., E. Stahl, S. Dudek, and M. Kreitman. 1998. Genetic variation within and among populations of *Arabidopsis thaliana*. Genetics **148**:1311–1323.

Brubaker, C. L., J. A. Koontz, and J. F. Wendel. 1993. Bidirectional cytoplasmic and nuclear introgression in the New World cottons, *Gossypium barbadense* and *G. hirsutum* (Malvaceae). Am. J. Bot. **80**:1203–1208.

Brubaker, C. L., and J. F. Wendel. 1994. Reevaluating the origin of domesticated cotton (*Gossypium hirsutum*; Malvaceae) using nuclear restriction fragment length polymorphisms (RFLPs). Am. J. Bot. **81**:1309–1326.

Chang, C., and E. Meyerowitz. 1986. Molecular cloning and DNA sequence of the *Arabidopsis thaliana* alcohol dehydrogenase gene. Proc. Natl. Acad. Sci. USA **83**:1408–1412.

Charlesworth, D., F.-L. Liu, and L. Zhang. 1998. The evolution of the alcohol dehydrogenase gene family by loss of introns in plants of the genus *Leavenworthia* (Brassicaceae). Mol. Biol. Evol. **15**:552–559.

Clegg, M. T. 1997. Plant genetic diversity and the struggle to measure selection. J. Hered. **88**:1–7.

Clegg, M. T., M. P. Cummings, and M. L. Durbin. 1997. The evolution of plant nuclear genes. Proc. Natl. Acad. Sci. USA **94**:7791–7798.

Cummings, M. P., and M. T. Clegg. 1998. Nucleotide sequence diversity at the alcohol dehydrogenase 1 locus in wild barley (*Hordeum vulgare* ssp. *spontaneum*): an evaluation of the background selection hypothesis. Proc. Natl. Acad. Sci. USA **95**:5637–5642.

Doebley, J. F. 1989. Isozymic evidence and the origin of crop plants. Pp. 165–186 *in* D. E. Soltis and P. S. Soltis, eds. Isozymes in plant biology. Dioscorides Press, Portland, Oreg.

———. 1992. Molecular systematics and crop evolution. Pp. 202–222 *in* P. S. Soltis, D. E. Soltis, and J. J. Doyle, eds. Molecular systematics of plants. Chapman and Hall, New York.

Drouin, G., and M. Moniz de Sá. 1997. Loss of introns in the pollen-specific actin gene subfamily members of potato and tomato. J. Mol. Evol. **45**:509–513.

Dvorák, J., M.-C. Luo, and Z.-L. Yang. 1998. Restriction fragment length polymorphism and divergence in the genomic regions of high and low recombination in self-fertilizing and cross-fertilizing Aegilops species. Genetics **148**: 423–434.

Endrizzi, J. D., E. L. Turcotte, and R. J. Kohel. 1985. Genetics, cytology, and evolution of *Gossypium*. Adv. Genet. **23**:271–375.

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.

Frugoli, J. A., M. A. McPeek, T. L. Thomas, and C. R. McClung. 1998. Intron loss and gain during evolution of the catalase gene family in angiosperms. Genetics **149**:355–365.

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

Gaut, B. S. 1998. Molecular clocks and nucleotide substitution rates in higher plants. Evol. Biol. **30**:93–120.

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

———. 1993*b*. Nucleotide polymorphism in the *Adh1* locus of pearl millet (*Pennisetum glaucum*) (Poaceae). Genetics **135**:1091–1097.

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

Hanfstingl, U., A. Berry, E. A. Kellogg, J. T. Costa III, W. Rüdiger, and F. M. Ausubel. 1994. Haplotypic divergence coupled with lack of diversity at the *Arabidopsis thaliana* alcohol dehydrogenase locus: roles for both balancing and directional selection? Genetics **138**:811–828.

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

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 of neutral molecular evolution based on nucleotide data. Genetics **116**:153–159.

Huttley, G. A., M. L. Durbin, D. E. Glover, and M. T. Clegg. 1997. Nucleotide polymorphism in the chalcone synthase-A locus and evolution of the chalcone synthase multigene family of common morning glory *Ipomoea purpurea*. Mol. Ecol. **6**:549–558.

INNAN, H., F. TAJIMA, R. TERAUCHI, and N. MIYASHITA. 1996. Intragenic recombination in the *Adh* locus of the wild plant *Arabidopsis thaliana.* Genetics **143**:1761–1770.

KAWABE, A., H. INNAN, R. TERAUCHI, and N. T. MIYASHITA. 1997. Nucleotide polymorphism in the acidic chitinase locus (*Chi*A) region of the wild plant *Arabidopsis thaliana.* Mol. Biol. Evol. **14**:1303–1315.

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.

KUMAR, S., K. TAMURA, and M. NEI. 1993. MEGA: molecular evolutionary genetics analysis. Version. 1.0. Pennsylvania State University, University Park.

LIU, F., L. ZHANG, and D. CHARLESWORTH. 1998. Genetic diversity in *Leavenworthia* populations with different inbreeding levels. Proc. R. Soc. Lond. B Biol. Sci. **265**:293–301.

MILLAR, A. A., and E. S. DENNIS. 1996. The alcohol dehydrogenase genes of cotton. Plant Mol. Biol. **31**:897–904.

MIYASHITA, N. T., H. INNAN, and R. TERAUCHI. 1996. Intra- and interspecific variation of the alcohol dehydrogenase locus region in wild plants *Arabis gemmifera* and *Arabidopsis thaliana.* Mol. Biol. Evol. **13**:433–436.

MORIYAMA, E. N., and J. R. POWELL. 1996. Intraspecific nuclear DNA variation in *Drosophila.* Mol. Biol. Evol. **13**: 261–277.

NEI, M. 1987. Molecular evolutionary genetics. Columbia University Press, New York.

NILES, G. A., and C. V. FEASTER. 1984. Breeding. Pp. 201–231 *in* R. J. KOHEL and C. F. LEWIS, eds. Cotton. American Society of Agronomy, Madison, Wis.

PERCY, R. G., and J. F. WENDEL. 1990. Allozyme evidence for the origin and diversification of *Gossypium barbadense* L. Theor. Appl. Genet. **79**:529–542.

POLLAK, E. 1987. On the theory of partially inbreeding finite populations. I. Partial selfing. Genetics **117**:353–360.

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

SAUER, J. 1967. Geographic reconnaissance of seashore vegetation along the Mexican Gulf Coast. Coastal Studies Series No. 21. Louisiana State University Press, Baton Rouge, La.

SEELANAN, T., A. SCHNABEL, and J. F. WENDEL. 1997. Congruence and consensus in the cotton tribe. Syst. Bot. **22**: 259–290.

SIMONSEN, K. L., G. A. CHURCHILL, and C. F. AQUADRO. 1995. Properties of statistical tests of neutrality for DNA polymorphism data. Genetics **141**:413–429.

SMALL, R. L., J. A. RYBURN, R. C. CRONN, T. SEELANAN, and J. F. WENDEL. 1998. The tortoise and the hare: choosing between noncoding plastome and nuclear *Adh* sequences for phylogeny reconstruction in a recently diverged plant group. Am. J. Bot. **85**:1301–1315.

STEPHENS, S. G. 1958. Salt water tolerance of seeds of *Gossypium* species as a possible factor in seed dispersal. Am. Nat. **92**:83–92.

STEWART, J. M. 1995. Potential for crop improvement with exotic germplasm and genetic engineering. Pp. 313–327 *in* G. A. CONSTABLE and N. W. FORRESTER, eds. Challenging the future: proceedings of the world cotton research conference-1. CSIRO, Melbourne.

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

———. 1993*a.* Simple methods for testing the molecular evolutionary clock hypothesis. Genetics **135**:599–607.

———. 1993*b.* Measurement of DNA polymorphism. Pp. 37–59 *in* N. TAKAHATA and A. G. CLARK, eds. Mechanisms of molecular evolution. Sinauer, Sunderland, Mass.

TERAUCHI, R., T. TERACHI, and N. T. MIYASHITA. 1997. DNA polymorphism at the *Pgi* locus of a wild yam, *Dioscorea tokoro.* Genetics **147**:1899–1914.

TRICK, M., E. S. DENNIS, K. J. R. EDWARDS, and W. J. PEACOCK. 1988. Molecular analysis of the alcohol dehydrogenase gene family of barley. Plant Mol. Biol. **11**:147–160.

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

WAYNE, M. L., and K. L. SIMONSEN. 1998. Statistical tests of neutrality in the age of weak selection. TREE **13**:236–240.

WENDEL, J. F. 1989. New World tetraploid cottons contain Old World cytoplasm. Proc. Natl. Acad. Sci. USA **86**:4132–4136.

———. 1995. Cotton. Pp. 358–366 *in* N. SIMMONDS and J. SMARTT, eds. Evolution of crop plants. Longman, London.

WENDEL, J. F., and V. A. ALBERT. 1992. Phylogenetics of the cotton genus (*Gossypium* L.): character-state weighted parsimony analysis of chloroplast DNA restriction site data and its systematic and biogeographic implications. Syst. Bot. **17**:115–143.

WENDEL, J. F., C. L. BRUBAKER, and A. E. PERCIVAL. 1992. Genetic diversity in *Gossypium hirsutum* and the origin of upland cotton. Am. J. Bot. **79**:1291–1310.

WENDEL, J. F., A. SCHNABEL, and T. SEELANAN. 1995. An unusual ribosomal DNA sequence from *Gossypium gossypioides* reveals ancient, cryptic, intergenomic introgression. Mol. Phylogenet. Evol. **4**:298–313.

WOLFE, K. H., W.-H. LI, and P. M. SHARP. 1987. Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. Proc. Natl. Acad. Sci. USA **84**:9054–9058.

WOLFE, K. H., P. M. SHARP, and W.-H. LI. 1989. Rates of synonymous substitution in plant nuclear genes. J. Mol. Evol. **29**:208–211.

PAMELA S. SOLTIS, reviewing editor

Accepted January 4, 1999