Differential Evolutionary Dynamics of Duplicated Paralogous *Adh* Loci in Allotetraploid Cotton (*Gossypium*)

Randall L. Small and Jonathan F. Wendel†*

*Department of Botany, The University of Tennessee, Knoxville; and †Department of Botany, Iowa State University

Levels and patterns of nucleotide diversity vary widely among lineages. Because allopolyploid species contain duplicated (homoeologous) genes, studies of nucleotide diversity at homoeologous loci may facilitate insight into the evolutionary dynamics of duplicated loci. In this study, we describe patterns of sequence diversity from an alcohol dehydrogenase homoeologous locus pair (*AdhC*) in allotetraploid cotton (*Gossypium*, Malvaceae). These data are compared with equivalent information from another homoeologous alcohol dehydrogenase gene pair (*AdhA*, Small, Ryburn, and Wendel 1999. Mol. Biol. Evol. **16**:491–501) which has an overall slower evolutionary rate than *AdhC.* As expected from the predicted correlation between nucleotide diversity and evolutionary rate, nucleotide diversity was higher for *AdhC* than for *AdhA*. In addition, nucleotide diversity is higher in the D-subgenome of allotetraploid cotton for *AdhC*, confirming earlier observations for *AdhA*. These observations indicate that for these two pairs of *Adh* loci, the null hypothesis of equivalent evolutionary dynamics for duplicated genes in allotetraploid cotton is rejected.

Introduction

Levels of diversity and patterns of substitution in genes are the footprints of the evolutionary processes that have shaped extant gene pools. Analyses of these patterns can provide insights into how the evolutionary process differs both between lineages and between loci within lineages (Clegg 1997; Clegg, Cummings, and Durbin 1997). In the absence of differential evolutionary pressures or genetic mechanisms, diversity is expected to be equivalent among loci, both in comparisons of orthologous genes between species and paralogous genes within species. Deviations from this expectation are the rule rather than the exception and may arise from myriad external and internal forces. Examples include variation in life history characteristics (e.g., self-pollination vs. outcrossing in plants [Liu, Zhang, and Charlesworth 1998; Savolainen et al. 2000]) and various forms of natural selection (e.g., primate ribonuclease genes [Zhang, Rosenberg, and Nei 1998]; gastropod toxin genes [Duda and Palumbi 1999]; plant self-incompatibility loci [Richman and Kohn 1999]; vertebrate MHC loci [Klein et al. 1998]; fungal mating type loci [May et al. 1999]). Additionally, it has been shown that nucleotide diversity is positively correlated with both evolutionary rates (Hudson, Kreitman, and Aguadé 1987) and recombination rates (Begun and Aquadro 1992). Finally, forces acting not on the gene of interest but on linked genes may also affect local levels and patterns of diversity because of background selection (Charlesworth, Morgan, and Charlesworth 1993; Cummings and Clegg 1998) or hitchhiking effects (Barton 1998; Przeworski, Charlesworth, and Wall 1999). Thus, differences in the levels and patterns of nucleotide diversity among loci and lineages may reflect numerous factors. To separate the effects of these various factors it is necessary to obtain data from multiple loci within a given phylogenetic framework.

Gossypium L. (Malvaceae) has become a useful model system for studying molecular evolution (Wendel, Schnabel, and Seelanan 1995; Cronn et al. 1996; Cronn, Small, and Wendel 1999; Small, Ryburn, and Wendel 1999; Small and Wendel 2000*a,* 2000*b*) and especially for studying the molecular evolutionary consequences of allopolyploidy (Wendel, Schnabel, and Seelanan 1995; Wendel et al. 1999; Wendel 2000; Liu et al. 2001). The phylogenetic relationships of the ca. 50 diploid and 5 allotetraploid species of *Gossypium* are well characterized (Wendel and Albert 1992; Seelanan, Schnabel, and Wendel 1997; Small et al. 1998; Wendel et al. 1999; Cronn et al. 2002). The five allotetraploid *Gossypium* species (designated AD-genome) diverged from a single recent allopolyploidization event (Wendel 1989; Small et al. 1998; Cronn, Small, and Wendel 1999), and the parental diploids are represented by the extant species *Gossypium herbaceum* L. (diploid A-genome) and *Gossypium raimondii* Ulbrich (diploid D-genome); thus the two component genomes of the allotetraploids are designated A- and D-subgenomes (or A′ and D′) to indicate their diploid origin. This well-understood organismal history facilitates the identification and comparison of orthologous and homoeologous loci (see e.g., Cronn and Wendel 1998; Small et al. 1998; Cronn, Small, and Wendel 1999; Small and Wendel 2000*a*).

A previous study (Small, Ryburn, and Wendel 1999) examined levels of nucleotide diversity for homoeologous *AdhA* loci in two allotetraploid species, *Gossypium hirsutum* L. and *Gossypium barbadense* L. Whereas that study revealed low diversity in both homoeologs, it also showed that the D-subgenome harbored greater nucleotide and allelic diversity than did the A-subgenome in both species. In concert with these data, a second study (Small et al. 1998) found that for a second alcohol dehydrogenase locus (*AdhC*), sequences from the D-subgenome homoeologs of all five allotetraploid species were evolving at a rate significantly greater than the rate in the A-subgenome homoeologs, again suggesting differential evolutionary pressures act-

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

Address for correspondence and reprints: Randall Small, Department of Botany, 437 Hesler Biology, The University of Tennessee, Knoxville, Tennessee 37996-1100. E-mail: rsmall@utk.edu.

Mol. Biol. Evol. 19(5):597–607. 2002
© 2002 by the Society for Molecular Biology and Evolution. ISSN: 0737-4038

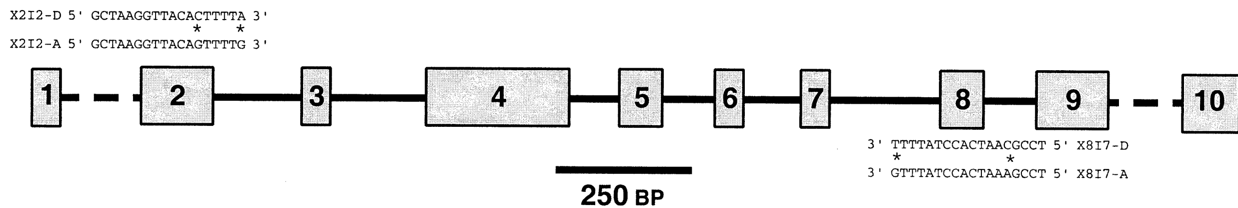

FIG. 1.—Diagrammatic representation of the *Gossypium AdhC* gene. Numbered boxes represent exons and intervening lines represent introns. The homoeolog-specific primers are shown in their approximate binding location; the nucleotide differences that confer homoeolog specificity are indicated by an *. Dashed lines for introns 1 and 9 denote that these regions of *AdhC* have not been sequenced. A 250-bp scale bar is shown for reference.

ing on the two subgenomes. Finally, in evaluating the relative rates for the entire *Adh* gene family in *Gossypium*, we found that *AdhC* has higher evolutionary rates at both silent and nonsynonymous sites than *AdhA* (Small and Wendel 2000*a*). Thus, evolutionary rates for *AdhA* are low, relative to those for *AdhC*. Because evolutionary rates and levels of nucleotide diversity are positively correlated (Hudson, Kreitman, and Aguadé 1987), these data predict that nucleotide diversity for *AdhC* should be higher than for *AdhA*. This, in turn, suggests that the observed increase of nucleotide and allelic diversity found in the D-subgenome of the allotetraploids for *AdhA* might similarly be elevated for *AdhC*. The purpose of this study then was to test these predictions for *AdhC*. Specifically we asked if: (1) nucleotide diversity is elevated for *AdhC* relative to *AdhA*, as predicted by the correlation between relative rates and nucleotide diversity; and (2) the pattern of higher diversity in the D-subgenome of the allotetraploids found for *AdhA* is also found for *AdhC*.

Materials and Methods

Plant Materials

Individual plants representing 22 accessions of *G. hirsutum* and six accessions of *G. barbadense* were included in this study. Each accession is representative of a wild-collected population or cultivar. These accessions are identical to those included in our previous study of *AdhA* (Small, Ryburn, and Wendel 1999), with the addition of a single *G. barbadense* accession (K101) which had been included in a previous study of *AdhC* (Small et al. 1998). Accessions were chosen to span the genetic and geographical variation encompassed by *G. hirsutum* and were originally selected based on the study of Brubaker and Wendel (1994), as described (Small, Ryburn, and Wendel 1999). *Gossypium* species as a general rule, and the cultivated allotetraploids in particular, are strongly selfing, intrapopulation variation is low, and heterozygosity is rare (Brubaker and Wendel 1994). Thus, our study was designed to maximize between-population, rather than within-population, sampling.

PCR Amplification and DNA Sequencing

To isolate *AdhC* sequences from specific duplicated genes in allotetraploid cotton, we designed two pairs of homoeolog-specific PCR amplification primers. Primer sequences were based on data from *AdhC* for all five allotetraploid species (Small et al. 1998) and were designed so that the final 3′ nucleotide of each primer, as well as one other nucleotide within the primer, were specific for either the A- or D-subgenome homoeolog. The forward primers span the exon 2-intron 2 boundary, whereas the reverse primers span the intron 7-exon 8 boundary (fig. 1). To achieve homoeolog-specific amplification, a two-step procedure was used. The first step involved a 10-μl PCR amplification using 0.5 μl of template DNA, 1× *Taq* buffer (Promega), 200 μM each dNTP, 1.5 mM $MgCl_2$, 0.2 μM each primer (either ADHCX2I2-D + ADHCX8I7-D to amplify the D-subgenome sequences or ADHCX2I2-A + ADHCX8I7-A to amplify the A-subgenome sequences). Cycling parameters used a touchdown approach (Don et al. 1991) that facilitates highly specific amplification. Initial annealing temperatures are set 5°C higher than the annealing temperature of the primers, so only amplification of the specific target is accomplished (in this case, the annealing temperature of the primers was 48–50°C, so the initial annealing temperature was set to 55°C). During the first 10 cycles, the annealing temperature is dropped by 0.5°C per cycle so that by the 11th cycle the programmed annealing temperature is down to the primer annealing temperature (50°C). An additional 15 cycles were then performed for a total of 25 cycles of 94°C for 1 min, 55–50°C for 1 min, and 72°C for 2 min, followed by a final 5 min 72°C extension step. The second step of the amplification process used 5 μl of the PCR product from the first step as a template for a second 25 μl PCR reaction with the same reaction components as above. Cycling conditions were 25 cycles of 94°C for 1 min, 50°C for 1 min, and 72°C for 2 min, followed by a final 5 min 72°C extension step. These PCR products were subjected to agarose gel electrophoresis, excised from the gel, and eluted from the gel using GeneClean (BIO 101).

Purified PCR products were sequenced directly either using the ThermoSequenase ^{33}P-radiolabeled terminator cycle sequencing kit (Amersham) followed by electrophoresis on a 5%–6% Long Ranger sequencing gel (FMC) or using the ABI Prism BigDye Terminator Cycle Sequencing Kit (Perkin-Elmer) followed by electrophoresis and detection on an ABI Prism 377 DNA sequencer at the Iowa State University DNA Sequencing and Synthesis Facility.

Because of the possibility that some mutations detected may be caused by nucleotide misincorporation by *Taq* polymerase during PCR, all singleton nucleotides were confirmed by reamplification and resequencing. In

all cases, the initial sequences inferred were corroborated by resequencing. In the case of heterozygous sequences, the amplification products were cloned into pGEM-T and sequenced as described previously (Small et al. 1998) to establish linkage relationships among polymorphic nucleotides.

Analyses

As in our previous study (Small, Ryburn, and Wendel 1999), we assumed that for each homoeolog both alleles were amplified. In a number of cases this assumption was validated by the presence of nucleotide polymorphism in the sequencing ladder and electropherograms, indicative of two products underlying the sequence (i.e., heterozygosity). The sequence uniformity detected for most accessions is assumed to be the result of homozygosity, the predominant condition in allotetraploid cottons (Wendel, Brubaker, and Percival 1992; Brubaker and Wendel 1994; Small, Ryburn, and Wendel 1999). Removing identical sequences inferred from homozygous individuals from the analyses has little quantitative effect on the results and does not change the qualitative conclusions.

The sequences generated fell into four subsets that were analyzed separately and in combination when appropriate. These data sets include the A-subgenome of *G. barbadense* (6 accessions, 12 alleles), the D-subgenome of *G. barbadense* (6 accessions, 12 alleles), the A-subgenome of *G. hirsutum* (22 accessions, 44 alleles), and the D-subgenome of *G. hirsutum* (22 accessions, 44 alleles).

Relationships among the haplotypes of the *AdhC* sequences from *G. hirsutum* and *G. barbadense* were inferred using the software TCS (Clement, Posada, and Crandall 2000), which implements a statistical parsimony approach to estimating gene genealogies (Posada and Crandall 2001). Genealogies were inferred separately for the A-subgenome sequences and D-subgenome sequences, but the *G. barbadense* and *G. hirsutum* sequences were analyzed together for each subgenome.

Descriptive statistics were calculated for each of the *AdhC* data sets. The two primary estimates of nucleotide diversity were π (Nei 1987, pp. 256–257) and θ_w (Watterson 1975), which estimate nucleotide diversity as the mean of all pairwise sequence differences and as an index of the number of polymorphic sites, respectively. A 95% confidence interval was calculated around θ_w using the coalescent simulation option of DnaSP v. 3.14 (Rozas and Rozas 1999). In addition, we calculated π separately for intron, synonymous, silent (intron + synonymous), and nonsynonymous sites.

A number of statistical tests have been proposed to evaluate whether or not the distribution of nucleotide polymorphism matches that predicted by neutral theory. We performed the tests of Tajima (1989), Fu and Li (1993), Hudson, Kreitman, and Aguadé (HKA 1987), and McDonald and Kreitman (MK 1991). Additionally, we explored the extent of recombination among sequences using the approach of Hudson and Kaplan (1985). This analysis infers the minimum number of recombination events within a collection of sequences using the four-gamete test in all pairwise comparisons of sequences. A number of these calculations were facilitated by the software DnaSP v. 3.14 (Rozas and Rozas 1999).

Results

AdhC Sequences

The portion of *AdhC* sequenced for this study is approximately 1.3 kb in length and included the majority of intron 2 through the majority of intron 7 (fig. 1). The sequences have been deposited in GenBank under accession numbers AF036569, AF036570, AF036575, AF036578, and AF403299–AF403367. Total aligned lengths ranged from 1,247 to 1,322 nt, including means of 709 intron sites, 139.5 synonymous sites, and 452.2 nonsynonymous sites. Two features of the data, both previously noted (Small et al. 1998), deserve mention with respect to the potential functionality of these genes. First, a 67-bp deletion is present in all sequences of the A-subgenome of *G. barbadense*. This deletion removes 7 nt from the 3′ end of exon 4, along with 60 nt from intron 4. Thus, *AdhC* may be a pseudogene in the A-subgenome of *G. barbadense*. Second, the first nucleotide of intron 6 in the D-subgenome of *G. hirsutum* is polymorphic for G (12 alleles) and A (32 alleles). Nuclear introns generally begin with the dinucleotide GT and end with the dinucleotide AG, which are important for intron splicing (Dibb 1993). Thus, *AdhC* in the D-subgenome of *G. hirsutum* is polymorphic for what may be an expression-altering mutation.

Gene Genealogies

Relationships among the *AdhC* sequences were inferred, and the resulting genealogies are depicted in figures 2 and 3. The A-subgenome network (fig. 2) reveals that *AdhC* sequences from *G. hirsutum* and *G. barbadense* are differentiated from each other by at least four mutations. Rooting this network with the diploid A-genome species *G. arboreum* places the root between the *G. barbadense* and *G. hirsutum* sequences. As discussed subsequently, no allelic diversity was detected in the A-subgenome of *G. barbadense*—all sequences were identical. Seven different haplotypes were recovered from the A-subgenome of *G. hirsutum*. Four of these haplotypes were found only in single homozygous individuals; the remaining three haplotypes were represented multiple times in the sample.

In contrast to the A-subgenome network, the D-subgenome network (fig. 3) reveals considerable haplotype diversity in *G. hirsutum*, where a total of 24 distinct haplotypes were observed, and *G. barbadense*, where three haplotypes were observed. Rooting this network with the diploid D-genome species *G. raimondii* divides the network into two neighborhoods, one above and one below the root (fig. 3). The neighborhood above the root includes eight different haplotypes. One of these haplotypes was found in both *G. hirsutum* and *G. barbadense*. Three of the six *G. barbadense* accessions sampled were homozygous for this shared haplotype,

FIG. 2.—Gene genealogy of *G. hirsutum* and *G. barbadense* A-subgenome *AdhC* sequences. Haplotypes observed in our study are represented by open circles (*G. hirsutum*) or open squares (*G. barbadense*); the number of times each haplotype was observed is indicated by the number inside the circle or square. Lines connecting the haplotypes represent a single nucleotide substitution with filled circles representing inferred mutational steps not observed. The position of the root (indicated by the arrow) is inferred by outgroup comparison with the diploid A-genome species *G. arboreum*. The dashed line indicates ambiguity in the connection of the root to the haplotypes because of missing data in the outgroup.

and it was also found in two of the *G. hirsutum* accessions, once in the homozygous condition and once as a part of a heterozygote. Five additional haplotypes found only in *G. hirsutum* accessions were also placed in this network neighborhood (fig. 3). The remaining *G. barbadense* sequences were represented by two different haplotypes, both also found in this neighborhood. The neighborhood below the root includes 18 different haplotypes, including the majority of the *G. hirsutum* alleles; no *G. barbadense* alleles were found in this neighborhood. A number of low-frequency haplotypes (represented only once in the sample) are found along the branch leading to this neighborhood, but the majority of the sequences, both in terms of numbers of different alleles and raw numbers of sequences, are found at the tip of the network. This includes one haplotype represented by eight sequences, three haplotypes represented by four sequences, and two haplotypes represented by two sequences (fig. 3). In addition, the recombination detected in these sequences (see subsequently) is indicated by the loop of four haplotypes in this neighborhood.

Nucleotide Polymorphism

For each data set we calculated two overall measures of nucleotide diversity, π and θ_w and a 95% confidence interval around θ_w (reported on a per base pair basis; table 1, fig. 4). π was also calculated separately for intron, synonymous, silent, and nonsynonymous sites. In addition, we determined the number of distinct alleles recovered in each data set and the minimum number of recombination events. The same values were

FIG. 3.—Gene genealogy of *G. hirsutum* and *G. barbadense* D-subgenome *AdhC* sequences. Haplotypes observed in our study are represented by open circles (*G. hirsutum*), open squares (*G. barbadense*), or a filled square (haplotype found in both *G. hirsutum* and *G. barbadense*); the number of times each haplotype was observed is indicated by the number inside the circle or square. Lines connecting the haplotypes represent a single nucleotide substitution with filled circles representing inferred mutational steps not observed. The position of the root (indicated by the arrow) is inferred by outgroup comparison with the diploid D-genome species *G. raimondii*. The loop connecting the four haplotypes depicts recombination relationships among those haplotypes.

calculated for the *AdhA* data (Small, Ryburn, and Wendel 1999) and are presented in table 1 for comparison.

Similar to the observations for *AdhA,* nucleotide diversity varies widely among genomes for *AdhC*. Values ranged from $\theta_w = 0/\pi = 0$ (*G. barbadense* A-sub-

Table 1
Estimates of Nucleotide Diversity Per Base Pair and Tests of Neutral Evolution for *AdhA* and *AdhC*. A′ and D′ Refer to the A- and D-Subgenomes of the Allotetraploids, Respectively

	N	θ_w	θ_w 95% CI	$\pi_{overall}$	$\pi_{silent (Synonymous + Introns)}$	$\pi_{Introns}$	$\pi_{Synonymous}$	$\pi_{non\text{-}synonymous}$	# Alleles	D[a]	D[b]	F[b]
AdhC												
G. hirsutum A′	44	0.00105	0.00017–0.00227	0.00085	0.00110	0.00131	0.00000	0.00039	7	−0.49	1.13	0.61
G. hirsutum D′	44	0.00522	0.00243–0.00939	0.00649	0.00558	0.00487	0.00926	0.00825	24	0.83	0.83	1.01
G. barbadense A′ . .	12	0.00000	—	0.00000	0.00000	0.00000	0.00000	0.00000	1	—	—	—
G. barbadense D′ . .	12	0.00200	0.00050–0.00451	0.00238	0.00272	0.00216	0.00563	0.00174	3	0.76	1.51	1.56
AdhA[c]												
G. hirsutum A′	44	0.00024	0.00000–0.00152	0.00050	0.00101	0.00000	0.00311	0.00000	2	1.47	0.55	0.94
G. hirsutum D′	44	0.00074	0.00016–0.00238	0.00123	0.00251	0.00243	0.00266	0.00000	4	1.42	0.91	1.24
G. barbadense A′ . .	10	0.00000	—	0.00000	0.00000	0.00000	0.00000	0.00000	1	—	—	—
G. barbadense D′ . .	10	0.00036	0.00001–0.00273	0.00036	0.00074	0.00000	0.00223	0.00000	2	0.62	0.74	0.67

[a] Tajima (1989); no results are significant.
[b] Fu and Li (1993); no results are significant.
[c] Data from Small, Ryburn and Wendel (1999).

genome) to $\theta_w = 0.00522/\pi = 0.00649$ (*G. hirsutum* D-subgenome). In all comparisons, both within and between *AdhA* and *AdhC*, the 95% confidence intervals for θ_w overlap, with the exception of the D-subgenome of *G. hirsutum* which has a significantly higher θ_w than all loci except *AdhA* and *AdhC* in the D-subgenome of *G. barbadense* (table 1, fig. 4).

However, nucleotide diversity is not equally distributed among site categories (intron, synonymous, silent, and nonsynonymous) in either *AdhA* or *AdhC*. As shown in table 1, all nucleotide diversity in *AdhA* is caused by silent polymorphism (either intron or synonymous sites); no nonsynonymous mutations were detected. In contrast, for *AdhC*, nonsynonymous diversity contributed a great deal to the observed variation. For *AdhC*, nonsynonymous diversity ranged from approximately half the overall diversity per site to actually exceeding overall diversity in one case (the D-subgenome of *G. hirsutum*).

Among putatively silent site categories (intron and synonymous sites), we also detected variation in nucleotide diversity (table 1). In almost all cases, synonymous diversity was greater than intron diversity. For *AdhA*, only the D-subgenome of *G. hirsutum* contained any intron diversity at all, and in this case synonymous diversity was only slightly higher than intron diversity. For both the A-subgenome of *G. hirsutum* and the D-subgenome of *G. barbadense*, all the diversity detected was at synonymous sites. For *AdhC*, synonymous diversity was approximately two times the intron diversity for the D-subgenomes of both *G. hirsutum* and *G. barbadense*. The A-subgenome of *G. hirsutum* provided the only exception to this trend, but in this case no synonymous diversity was detected. This pattern of higher diversity and divergence at synonymous sites relative to intron sites has been noted previously both in plants and Drosophila (Moriyama and Powell 1996; Charlesworth and Charlesworth 1998; Vieira and Charlesworth 2001) and is presumably caused by greater selective constraints on sites important for intron structure relative to synonymous changes in coding regions.

Nucleotide diversity values of *AdhC* for a given genome are consistently higher than for *AdhA* (except for the A-subgenome of *G. barbadense* which was monomorphic for both *AdhA* and *AdhC*). Values for θ_w were 4.4–7.1 times higher for *AdhC* than *AdhA*, whereas π values were 1.7–6.6 times higher. Despite this elevation of diversity in *AdhC* relative to *AdhA*, these values are still low compared with plant nuclear genes in general. For example, the highest estimate of π in *Gossypium* is 0.00649 (*G. hirsutum* D-subgenome *AdhC*), whereas π in *Arabidopsis thaliana* has a mean of 0.00665 and ranges from 0.00300 to 0.01040 for five nuclear genes (*Adh*, [Innan et al. 1996]; *CAL* [Purugganan and Suddith 1998]; *ChiA* [Kawabe et al. 1997]; *ChiB* [Kawabe and Miyashita 1999]; *CHI* [Kuittinen and Aguadé 2000]; *FAH1* [Aguadé 2001]; *F3H* [Aguadé 2001]; and *RPS2* [Caicedo, Schaal, and Kunkel 1999]).

Recombination Indices and Neutrality Tests

The minimum number of recombination events per data set was inferred using the method of Hudson and

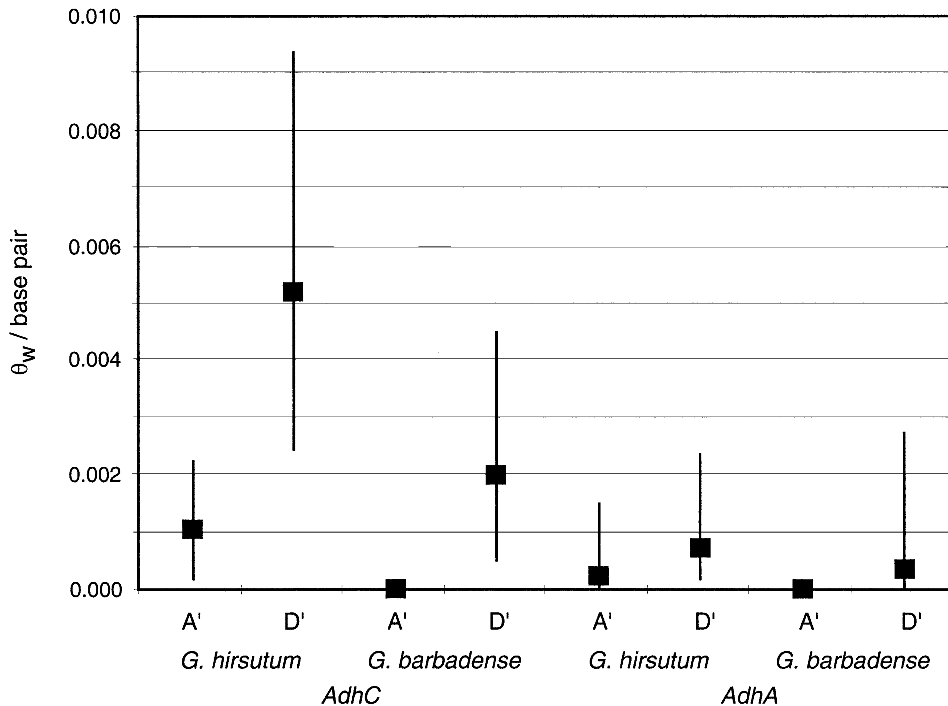

FIG. 4.—Estimates of θ_w for each gene-species-subgenome sampled are indicated by the filled squares. The 95% confidence interval inferred for each estimate is indicated by the line.

Kaplan (1985). Recombination was detected only in the *AdhC G. hirsutum* D-subgenome data set. As noted above, these recombination events are restricted to a set of four closely related haplotypes (fig. 3). The tests of neutral evolution of Tajima (1989) and Fu and Li (1993) were performed for each data set, including subsets of each data set (introns and exons). None of these tests revealed significant departures from neutral expectations for any data set. We also performed the HKA test (Hudson, Kreitman, and Aguadé 1987) and the MK test (McDonald and Kreitman 1991). For the HKA test, the data sets were partitioned as follows: the intraspecific comparison was between *AdhC* from the A-subgenome of *G. hirsutum* and *AdhC* from the D-subgenome of *G. hirsutum*; the interspecific comparison was provided by the *AdhC* sequences from the A- and D-subgenomes of *G. barbadense*. This test did not reveal any departure from neutrality ($\chi^2 = 0.84$, $P = 0.36$). The MK test was performed by tabulating numbers of fixed and polymorphic synonymous and nonsynonymous substitutions in exons for all four data sets and performing a *G*-test (with Williams correction) of independence. The MK test did reveal a significant departure from neutrality ($G = 6.924$, $P = 0.0085$, fig. 5) caused by an excess of polymorphic replacement substitutions.

Discussion

Levels and patterns of nucleotide diversity in *Adh* genes of allotetraploid *Gossypium* species differ markedly in several ways. First, the gene genealogies of *AdhC* in the A- and D-subgenomes differ in their patterns of haplotype distribution. Second, comparisons between homoeologous locus pairs (*AdhA* vs. *AdhC*) reveal that *AdhC* has consistently higher nucleotide and allelic diversity than *AdhA*. Finally, comparisons between subgenomes show that the D-subgenomes of *G. hirsutum* and *G. barbadense* harbor greater nucleotide and allelic diversity than the A-subgenomes. The sum of these observations indicates that the evolutionary forces that have shaped the nucleotide diversity at these loci have differed, both between loci and between subgenomes.

Gene Genealogies and Coalescence

AdhC gene genealogies were constructed separately for the A- and D-subgenome sequences of *G. hirsutum* and *G. barbadense* (figs. 2 and 3). These genealogies reveal different patterns of haplotype distribution in the two subgenomes. The A-subgenome sequences reflect a simple underlying pattern (fig. 2): *G. hirsutum* and *G. barbadense* sequences are separated on the genealogy by at least four substitutions. If this tree is rooted with an *AdhC* sequence from an A-genome diploid species (*G. arboreum*), the root falls on the branch separating the *G. hirsutum* and *G. barbadense* sequences; i.e., *AdhC* alleles coalesce within species. All *G. barbadense* sequences fell into a single haplotype. *Gossypium hirsutum* sequences were represented by seven different haplotypes; three of these were at intermediate frequency and four were found as homozygotes in single individuals. No recombination was detected among these sequences.

The pattern depicted in the genealogy of the D-subgenome sequences, on the other hand, is more complex (fig. 3). If this tree is rooted with an *AdhC* sequence from a D-genome diploid species (*G. raimondii*), the

```
                               Nucleotide           111111222222344455555555
                               Position      2467778012289146778612612567889
                                             4475785378996967690297164266361
                               -------------------------------------------------
G. hirsutum    A-subgenome     pfx_1A        TGGGCGGCCAGTGGCCTCGGCGCCTCCTCGA
                               TX1_1A        ...............................
                               TX6_1A        ...............................
                               TX21_1A       ............................G..
                               TX44_1A       ...............................
                               TX51_1A       ...............................
                               TX93_1A       ...............................
                               TX94_1A       .....................G.........
                               TX98_1A       ...............................
                               TX111_1A      ...............................
                               TX116_1A      ...............................
                               TX119_1A      ...............................
                               TX166_1A      ...............................
                               TX188_1A      ...............................
                               TX192_1A      ...............................
                               TX367_1A      ...............................
                               TX481_1A      ...............................
                               TX706_1A      ...............................
                               TX766_1A      ...............................
                               PAY_1A        ...............................
                               DEL_1A        ...............................
                               BR115_1A      ...............................
                               -------------------------------------------------
G. barbadense  A-subgenome     K101_1A       ...............................
                               PIMA_1A       ...............................
                               B106_1A       ...............................
                               B250_1A       ...............................
                               B444_1A       ...............................
                               B559_1A       ...............................
                               -------------------------------------------------
G. hirsutum    D-subgenome     pfx_1D        C.AATAC...AC.TTTC.AAT...CT.C.AG
                               TX1_1D        CA.ATAC...AC.TTTC.AA.A.TCT.C...
                               TX1_2D        CA.ATAC...A..TTTC....A.T.T.C...
                               TX6_1D        C.AATAC...AC.TTTC.AAT...CTTC.AG
                               TX21_1D       CAAATAC...AC.TTTC.AA.A.T.T.C...
                               TX21_2D       CA.ATACG..A..TTTC....A.T.T.C...
                               TX44_1D       CAAATAC..GAC.TTTC.AA.A.T.T.C...
                               TX44_2D       CA.ATAC..GA..TTTC....A.T.T.C...
                               TX51_1D       CA.ATAC...A..TTTC....A.T.T.C...
                               TX93_1D       C.AATAC...AC.TTTC.AAT...CT.C..G
                               TX93_2D       C.AATAC...AC.TTTC.AAT...CTTC..G
                               TX94_1D       C.AATAC...AC.TTTC.AAT...CTTC..G
                               TX94_2D       CA.ATAC...A..TTTC....A.T.T.C...
                               TX98_1D       C.AATAC...AC.TTTC.AAT...CTTC..G
                               TX98_2D       CA.ATAC...A..TTTC....A.T.T.C...
                               TX111_1D      C.AATAC...AC.TTTC.AAT...CTTC..G
                               TX111_2D      CA.ATAC...A..TTTC....A.T.T.C...
                               TX116_1D      C.AATAC...AC.TTTC.AAT...CT.C.AG
                               TX119_1D      C.AATAC...AC.TTTC.AAT...CTTC..G
                               TX119_2D      CAAATAC...AC.TTTC.AATA.TCT.C...
                               TX166_1D      C.AATAC...AC.TTTC.AAT...CTTC..G
                               TX166_2D      C.AATAC...AC.TTTC.AAT...CT.C..G
                               TX188_1D      C.AATAC...AC.TTTC.AAT...CTTC..G
                               TX188_2D      C.AATAC...AC.TTTC.AAT..TCT.C..G
                               TX192_1D      C.AATAC...AC.TTTC.AAT...CTTC..G
                               TX367_1D      C.AATAC.T.ACCTTTC.AAT..TCT.C.AG
                               TX367_2D      CA.ATAC.T.ACCTTTC.AAT..TCT.C...
                               TX481_1D      C.AATAC...AC.TTTC.AAT...CTTC..G
                               TX706_1D      CAAATAC..GAC.TTTC.AAT...CT.C.AG
                               TX706_2D      CA.ATAC..GAC.TTTC.AAT...CT.C.AG
                               TX766_1D      C.AATAC...AC.TTTC.AAT...CTTC..G
                               TX766_2D      C.AATAC...AC.TTTC.AAT...CT.C..G
                               PAY_1D        C.AATAC...AC.TTTC.AAT...CTTC..G
                               PAY_2D        C.AATAC...AC.TTTC.AAT...CT.C..G
                               DEL_1D        C.AATAC...AC.TTTC.AAT...CTTC..G
                               DEL_2D        C.AATAC...AC.TTTC.AAT..T.T.C..G
                               BR115_1D      C.AATAC...AC.TTTC.AAT...CTTC..G
                               BR115_2D      C.AATAC...AC.TTTC.AAT...CT.C..G
                               -------------------------------------------------
G. barbadense  D-subgenome     K101_1D       CA.ATAC...A..TTTC....A.T.T.C...
                               PIMA_1D       CA.ATAC...A..TTTC....A.T.T.C...
                               B106_1D       CA.ATAC...A..TTTC....A.T.T.C...
                               B250_1D       C..ATAC...A..TTTCT...A...T.C...
                               B444_1D       C..ATAC..GA..TTTC....A...T.C...
                               B559_1D       CA.ATAC...A..TTTC....A.T.T.C...
                               -------------------------------------------------
                               SYN/REP       SRRSRSRRSRSSRRSSSSRRRRRSRSRSRRS
                               FIXED/POLY    FPPFFFFFPPPFPPFFFFPPPPPPPPPFPFPPP
```

	Fixed	Polymorphic
Replacement	3	14
Synonymous	9	5

G-test $P = 0.0085$

Site	Amino Acid Change
44	Gly - Glu
67	Glu - Lys
77	Thr - Met
85	Glu - Pro
103	Ile - Cys
128	Lys - Arg
196	Gly - Arg
219	Gln - His
280	Glu - Lys
362	Gly - Glu
419	Ser - Phe
427	Val - Ile
461	Ala - Gly
524	Phe - Ser
566	Pro - Leu
583	Pro - Ala
586	Val - Ile

FIG. 5.—Data and results for the McDonald-Kreitman test. The complete data set was trimmed to include only exon sequences, and only those sites that are polymorphic within that data set are shown. The nucleotide position numbers reflect the positions within the trimmed data set. All sequences are shown relative to the top sequence (pfx_1A) with a "." indicating identity with the reference sequence. The status of each polymorphism (synonymous vs. replacement; fixed vs. polymorphic) is indicated at the bottom of each column. These values are compiled into a 2 × 2 contingency table and analyzed with a *G*-test (with Williams correction). The resulting $P = 0.0085$ indicates a significant departure from neutral expectations caused by an excess of polymorphic replacement substitutions. The amino acid changes are shown for each of the replacement substitutions.

genealogy is divided into two neighborhoods. Three different *G. barbadense* haplotypes were recovered, all of which fall into the neighborhood above the root. The majority (8/12) of the *G. barbadense* sequences fell into a single haplotype; two additional haplotypes were observed as homozygotes in single individuals. The most frequent *G. barbadense* haplotype was also observed in two *G. hirsutum* accessions, once as a homozygote and once as part of a heterozygote. A number of additional *G. hirsutum* haplotypes are also observed in this neighborhood, each being one to four substitutions different from the most frequent haplotype in this neighborhood. The remaining *G. hirsutum* sequences are found in the neighborhood below the root. A long branch connects the two neighborhoods, and most haplotypes are found near the ends of these neighborhoods, although a few low-frequency *G. hirsutum* haplotypes are found along the branch.

D-subgenome *AdhC* sequences from *G. hirsutum* do not coalesce within the species; in fact, a number of haplotypes found in *G. hirsutum* are more closely related to *G. barbadense* sequences than they are to *G. hirsutum* sequences of the other neighborhood, and one haplotype is shared by *G. barbadense* and *G. hirsutum*. One explanation for the transspecies polymorphism observed is that it is caused by the inheritance of ancient polymorphism(s) from the common ancestor of *G. hirsutum* and *G. barbadense*. An alternative explanation is that these sequences have been introgressed from *G. barbadense* into *G. hirsutum*, a phenomenon previously observed between these two species (Brubaker, Koontz, and Wendel 1993). It may be noteworthy in this respect that the *G. hirsutum* accessions with *G. barbadense*-like haplotypes all occur in the southern Mexican states of Chiapas and Guerrero or in neighboring Guatemala, the region of sympatry between *G. hirsutum* and *G. barbadense*. The pattern of introgression described by Brubaker, Koontz, and Wendel (1993) is consistent with our data, in that they detected introgression from *G. barbadense* into *G. hirsutum* primarily in wild or feral populations in the region of sympatry. Introgression from *G. hirsutum* into *G. barbadense*, however, was generally restricted to modern cultivars.

However, the transspecies polymorphism is observed only in the D-subgenome sequences, not in the A-subgenome sequences. Introgression would not be expected to be restricted to a single locus unless strong selection was acting to promote introgression in the D-subgenome or to prevent it in the A-subgenome. No evidence of such selection pressure has been demonstrated.

Importantly, regardless of the ultimate source of these *G. barbadense*-like alleles, their impact on the patterns of diversity is not overwhelming. If these sequences are removed from the analyses, π in the D-subgenome of *G. hirsutum* drops from 0.00649 to 0.00443, and θ_w drops from 0.00522 to 0.00507—these values are still well above those observed in other *Gossypium* species or subgenomes. Additionally, the results of the MK test are still significant if the *G. barbadense*-like alleles are excluded.

Comparative Evolutionary Dynamics of *AdhA* versus *AdhC*

Our previous study demonstrated that nucleotide substitution rates are higher for *AdhC* than for *AdhA* at both silent and nonsynonymous sites (Small and Wendel 2000*a*). Neutral theory predicts that evolutionary rates and nucleotide diversity will be positively correlated (Hudson, Kreitman, and Aguadé 1987), suggesting that nucleotide diversity should be higher for *AdhC* than for *AdhA*. That expectation is confirmed by our data, where on a per genome basis nucleotide diversity is higher for *AdhC* in every comparison. Furthermore, allelic diversity is consistently higher for *AdhC* than *AdhA* with 26 unique haplotypes recovered for *AdhC* (24 in the D-subgenome of *G. hirsutum*) as opposed to six haplotypes for *AdhA* (a maximum of four in any single genome, again in the D-subgenome of *G. hirsutum*). In addition, haplotype diversity is more widely dispersed on the gene genealogy in the D-subgenome than in the A-subgenome. *Gossypium hirsutum* D-subgenome sequences differ by up to 27 nucleotide substitutions (fig. 3), whereas the most divergent A-subgenome sequences differ by only four nucleotide substitutions (fig. 2).

In addition to the higher overall diversity of *AdhC* relative to *AdhA*, the patterns of silent and nonsynonymous diversity for the two loci are different. Specifically, no nonsynonymous diversity was detected for the *AdhA* genes, whereas nonsynonymous diversity accounts for a significant portion of the diversity at *AdhC* (table 1). Variation in silent versus nonsynonymous evolutionary rates has previously been described in plant genomes. For example, Gaut (1998) examined rate variation among nine nuclear genes for a rice-maize comparison and found that synonymous rates varied over a 2.4-fold range, and nonsynonymous rates varied over a 10-fold range. More relevant to the present study, five loci of the *Gossypium Adh* gene family have synonymous rates that vary over a 2.9-fold range and nonsynonymous rates that vary over a 3.3-fold range (Small and Wendel 2000*a*). The source of this variation in relative rates may be caused by either genomic processes that differentially affect the two loci, differential selective pressures on the two loci, or a combination of these factors. Recent evidence from extensive analyses of mammalian genomes suggests that evolutionary rates vary by genomic region, with genes from the same region showing similar synonymous rates (Matassi, Sharp, and Gautier 1999). Alternatively, different selection pressures on the two loci may be responsible for the observed differences in silent and nonsynonymous diversity. Support for this hypothesis is provided by the results of the MK test, which reveals that the patterns of synonymous and replacement substitutions at *AdhC* are not in accordance with neutral expectations. Specifically, there is an excess of polymorphic replacement substitutions, the majority of which (12/14) are polymorphic in the D-subgenomes of *G. hirsutum* or *G. barbadense* (or both). This observation contrasts with the lack of replacement substitutions in *AdhA*, suggesting differential selective pressures on *AdhA* and *AdhC*, ei-

ther purifying selection on *AdhA*, relaxed selection on *AdhC*, or a combination of the two. The lack of significant results for the Tajima or Fu and Li neutrality tests suggests that there is no disruption of the pattern of a neutral array of nucleotide substitutions, although the power of these tests is notoriously low (Simonsen, Churchill, and Aquadro 1995), and the effect of deviations from the tests assumptions (e.g., random mating—*Gossypium* species are strongly selfing) is unknown.

Thus, our observation of consistently greater nucleotide diversity and elevated nonsynonymous substitution rates in comparing *AdhA* and *AdhC* may be accounted for either by differential genomic context of the two genes, differential selective pressures on the two genes, or a combination of the two phenomena. Evidence is presented for differential selective pressures; the influence of genomic context, however, cannot be evaluated until similar data are available for genes in the same genomic context as *AdhA* and *AdhC*.

Comparative Evolutionary Dynamics of A-subgenome versus D-subgenome Sequences

Whereas our data clearly show that evolutionary dynamics differ between loci (*AdhA* vs. *AdhC*), the data also suggest differential patterns of evolution between sequences from the A- and D-subgenomes of allotetraploid *Gossypium*. In all pairwise comparisons of nucleotide diversity between subgenomes within a species (e.g., *G. hirsutum* A-subgenome vs. *G. hirsutum* D-subgenome for *AdhC*), nucleotide diversity is consistently higher in the D-subgenome (table 1, fig. 4). Likewise, the number of haplotypes recovered in each data set is consistently higher in the D-subgenome, with ratios ranging from 2:1 to 3.4:1 (table 1). Further, relative rate tests for *AdhC* have shown that the D-subgenome sequences are evolving at a significantly faster rate than A-subgenome sequences (Small et al. 1998; Small and Wendel 2000*a*). Finally, the excess of polymorphic replacement substitutions at *AdhC*, as evidenced both by the MK test and the high nonsynonymous diversity in the D-subgenome of *G. hirsutum*, suggests relaxed selection on the D-subgenomes of *G. hirsutum* and *G. barbadense*, at least for *AdhC*.

The genetic redundancy created by allopolyploidy or the large *Adh* gene family in *Gossypium* (at least seven loci in the diploid species [Small and Wendel 2000*a*]) may have allowed relaxed selection in the D-subgenome, whereas purifying selection maintained a narrower array of A-subgenome sequences. This hypothesis is consistent, with respect to both *AdhA* and *AdhC*, with the elevated evolutionary rate of the D-subgenome sequences over the A-subgenome sequences (Small et al. 1998) and the higher diversity in the D-subgenomes relative to the A-subgenomes. In addition, the presence of an intron-splice site mutation segregating in the *G. hirsutum AdhC* gene further suggests that, at least in *G. hirsutum*, this locus may be in the process of becoming a pseudogene.

Collectively, these observations might suggest an overall rate acceleration in the D-subgenome relative to the A-subgenome of allotetraploid *Gossypium*. Evolutionary rate analyses of 14 other nuclear loci, however, fail to reinforce this conclusion (Cronn, Small, and Wendel 1999). The apparent contradiction between the results of Cronn, Small, and Wendel (1999) and the present study indicates that either the *Adh* data are unusual, attributed perhaps to stochastic factors; the power of the relative rate tests alone are insufficient to detect subtle inequalities between subgenomes (as opposed to a combination of relative rate and nucleotide diversity analyses); or that something specific to these *Adh* loci promotes greater diversity in one of the two cotton subgenomes. We are currently unable to discriminate between these possibilities. Expression data and comparable studies of additional duplicated loci may provide critical clues in unraveling this conundrum.

Acknowledgments

We thank Julie Ryburn for technical assistance during the course of this investigation; Brandon Gaut and two anonymous reviewers provided helpful comments on an earlier draft of this paper; and funding was provided by the National Science Foundation to J.F.W.

LITERATURE CITED

AGUADÉ, M. 2001. Nucleotide sequence variation at two genes of the phenylpropanoid pathway, the *FAH1* and *F3H* genes, in *Arabidopsis thaliana*. Mol. Biol. Evol. **18**:1–9.

BARTON, N. H. 1998. The effect of hitch-hiking on neutral genealogies. Genet. Res. **72**:123–133.

BEGUN, D. J., and C. F. AQUADRO. 1992. Levels of naturally occuring DNA polymorphism correlate with recombination rates in *Drosophila melanogaster*. Nature **356**:519–520.

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.

CAICEDO, A. L., B. A. SCHAAL, and B. N. KUNKEL. 1999. Diversity and molecular evolution of the *RPS2* resistance gene in *Arabidopsis thaliana*. Proc. Natl. Acad. Sci. USA **96**:302–306.

CHARLESWORTH, B., M. T. MORGAN, and D. CHARLESWORTH. 1993. The effect of deleterious mutations on neutral molecular variation. Genetics **134**:1289–1303.

CHARLESWORTH, D., and B. CHARLESWORTH. 1998. Sequence variation: looking for effects of genetic linkage. Curr. Biol. **8**:R658–R661.

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.

CLEMENT, M., D. POSADA, and K. A. CRANDALL. 2000. TCS: a computer program to estimate gene genealogies. Mol. Ecol. **9**:1657–1659.

CRONN, R., and J. F. WENDEL. 1998. Simple methods for isolating homoeologous loci from allopolyploid genomes. Genome **41**:756–762.

CRONN, R. C., R. L. SMALL, T. HASELKORN, and J. F. WENDEL. 2002. Rapid diversification of the cotton genus (*Gossypium*: Malvaceae) revealed by analysis of sixteen nuclear and chloroplast genes. Am. J. Bot. (in press).

CRONN, R. C., R. L. SMALL, and J. F. WENDEL. 1999. Duplicated genes evolve independently after polyploid formation in cotton. Proc. Natl. Acad. Sci. USA **96**:14,406–14,411.

CRONN, R. C., X. ZHAO, A. H. PATERSON, and J. F. WENDEL. 1996. Polymorphism and concerted evolution in a tandemly repeated gene family: 5S ribosomal DNA in diploid and allopolyploid cottons. J. Mol. Evol. **42**:685–705.

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.

DIBB, N. J. 1993. Why do genes have introns? FEBS Lett. **325**: 135–139.

DON, R. H., P. T. COX, B. J. WAINWRIGHT, K. BAKER, and J. S. MATTICK. 1991. 'Touchdown' PCR to circumvent spurious priming during gene amplification. Nucleic Acids Res. **19**:4008.

DUDA, T. F., and S. R. PALUMBI. 1999. Molecular genetics of ecological diversification: duplication and rapid evolution of toxin genes of the venomous gastropod *Conus*. Proc. Natl. Acad. Sci. USA **96**:6820–6823.

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.

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. K. KREITMAN, and M. AGUADÉ. 1987. A test of neutral molecular evolution based on nucleotide data. Genetics **116**:153–159.

INNAN, H., F. TAJIMA, R. TERAUCHI, and N. T. 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 (*ChiA*) region of the wild plant *Arabidopsis thaliana*. Mol. Biol. Evol. **14**:1303–1315.

KAWABE, A., and N. T. MIYASHITA. 1999. DNA variation in the basic chitinase locus (*ChiB*) region of the wild plant *Arabidopsis thaliana*. Genetics **153**:1445–1453.

KLEIN, J., A. SATO, S. NAGL, and C. O'HUIGIN. 1998. Molecular trans-species polymorphism. Ann. Rev. Ecol. Syst. **29**: 1–21.

KUITTINEN, H., and M. AGUADÉ. 2000. Nucleotide variation at the *CHALCONE ISOMERASE* locus in *Arabidopsis thaliana*. Genetics **155**:863–872.

LIU, B., C. L. BRUBAKER, G. MERGEAI, R. C. CRONN, and J. F. WENDEL. 2001. Polyploid formation in cotton is not accompanied by rapid genomic changes. Genome **44**:321–330.

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

MATASSI, G., P. M. SHARP, and C. GAUTIER. 1999. Chromosomal location effects on gene sequence evolution in mammals. Curr. Biol. **9**:786–791.

MAY, G., F. SHAW, H. BADRANE, and X. VEKEMANS. 1999. The signature of balancing selection: fungal mating compatibility gene evolution. Proc. Natl. Acad. Sci. USA **96**: 9172–9177.

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

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.

POSADA, D., and K. A. CRANDALL. 2001. Intraspecific gene genealogies: trees grafting into networks. TREE **16**:37–45.

PRZEWORSKI, M., B. CHARLESWORTH, and J. D. WALL. 1999. Genealogies and weak purifying selection. Mol. Biol. Evol. **16**:246–252.

PURUGGANAN, M. D., and J. I. SUDDITH. 1998. Molecular population genetics of the Arabidopsis CAULIFLOWER regulatory gene: nonneutral evolution and naturally occurring variation in floral homeotic function. Proc. Natl. Acad. Sci. USA **95**:8130–8134.

RICHMAN, A. D., and J. R. KOHN. 1999. Self-incompatibility alleles from Physalis: implications for historical inference from balanced genetic polymorphisms. Proc. Natl. Acad. Sci. USA **96**:168–172.

ROZAS, J., and R. ROZAS. 1999. DnaSP version 3: an integrated program for molecular population genetics and molecular evolution analysis. Bioinformatics **15**:174–175.

SAVOLAINEN, O., C. H. LANGLEY, B. P. LAZZARO, and H. FREVILLE. 2000. Contrasting patterns of nucleotide polymorphism at the alcohol dehydrogenase locus in the outcrossing *Arabidopsis lyrata* and the selfing *Arabidopsis thaliana*. Mol. Biol. Evol. **17**:645–655.

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

SIMONSEN, K. L., G. A. CHURCHILL, and C. A. 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 phylogenetic reconstruction in a recently diverged plant group. Am. J. Bot. **85**:1301–1315.

SMALL, R. L., J. A. RYBURN, and J. F. WENDEL. 1999. Low levels of nucleotide diversity at homoeologous *Adh* loci in allotetraploid cotton (*Gossypium* L.). Mol. Biol. Evol. **16**: 491–501.

SMALL, R. L., and J. F. WENDEL. 2000*a*. Copy number lability and evolutionary dynamics of the *Adh* gene family in diploid and tetraploid cotton (*Gossypium*). Genetics **155**:1913–1926.

———. 2000*b*. Phylogeny, duplication, and intraspecific variation of *Adh* sequences in New World diploid cottons (*Gossypium*, Malvaceae). Mol. Phylogenet. Evol. **16**:73–84.

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

VIEIRA, C. P., and D. CHARLESWORTH. 2001. Low diversity and divergence in the *fil1* gene family of *Antirrhinum* (Scrophulariaceae). J. Mol. Evol. **52**:171–181.

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

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

———. 2000. Genome evolution in polyploids. Plant Mol. Biol. **42**:225–249.

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. Bidirectional interlocus concerted evolution following allopolyploid speciation in cotton (*Gossypium*). Proc. Natl. Acad. Sci. USA **92**:280–284.

WENDEL, J. F., R. L. SMALL, R. C. CRONN, and C. L. BRUBAKER. 1999. Genes, jeans, and genomes: reconstructing the history of cotton *in* L. W. D. van Raamsdonk and J. C. M. den Nijs, eds. Plant evolution in man-made habitats. Proceedings of VIIth International Symposium of the International Organization of Plant Biosystematists. Hugo de Vries Laboratory, Amsterdam.

ZHANG, J., H. F. ROSENBERG, and M. NEI. 1998. Positive Darwinian selection after gene duplication in primate ribonuclease genes. Proc. Natl. Acad. Sci. USA **95**:3708–3713.

BRANDON GAUT, reviewing editor

Accepted October 19, 2001