Molecular Phylogenetics and Evolution
Vol. 16, No. 1, July, pp. 73–84, 2000
doi:10.1006/mpev.1999.0750, available online at http://www.idealibrary.com on IDEAL®

# Phylogeny, Duplication, and Intraspecific Variation of *Adh* Sequences in New World Diploid Cottons (*Gossypium* L., Malvaceae)

Randall L. Small[1] and Jonathan F. Wendel

*Department of Botany, Iowa State University, Ames, Iowa 50011*

Received July 23, 1999; revised October 13, 1999

**The 13 "D-genome" cotton species are a monophyletic assemblage of morphologically diverse diploids that inhabit arid to semiarid regions in Mexico, with 1 disjunct species each in Peru and the Galapagos Islands and 1 species whose range extends northward into Arizona. While these species lack commercially significant fiber (i.e., cotton), they are important in that they represent one of the parental genomes of the cultivated tetraploid cottons. To assess phylogenetic relationships among these species, we sequenced and analyzed a region of a nuclear-encoded alcohol dehydrogenase gene (*AdhA*). Phylogenetic analysis resulted in a topology that is generally consistent with current taxonomic alignment of the species, although the phylogeny based on *AdhA* sequences conflicts with those inferred from cpDNA and ITS data sets, most notably in the position of the anomalous species *Gossypium gossypioides*. In one lineage, we detected both gene duplication and sequence polymorphisms that transcend species boundaries; sequences in this lineage formed a monophyletic clade, yet no taxon within the clade contained a monophyletic collection of sequences. Potential explanations for this latter phenomenon, including gene duplication, gene flow, and lineage sorting, are discussed.** 

## INTRODUCTION

The cotton genus, *Gossypium* L. (Malvaceae), includes approximately 50 species distributed throughout the tropical and subtropical regions of the world. While the primary cultivated cotton, *G. hirsutum* L., is a tetraploid, the majority of the wild species are diploid ($2n = 2x = 26$; Endrizzi *et al.*, 1985; Wendel, 1995). The diploid species have been divided into cytologically based genome groups (reviewed by Endrizzi *et al.*, 1985; Stewart, 1995) and exist in three primary centers of diversity: the African–Asian species (A-, B-, E-, and F-genomes), the Australian species (C-, G-, and K-genomes), and the New World species (D-genome). The 13 New World diploid D-genome species are the focus of this report.

The D-genome diploid species are distributed throughout Mexico, with one disjunct species each in Peru and the Galapagos Islands and one whose range extends northward into Arizona. These species are classified in *Gossypium* subgenus *Houzingenia* Fryxell (Fryxell, 1992) which is divided into two sections and six subsections (Table 1). Given their potential importance in germplasm improvement and understanding of the origin of the commercially important tetraploid species, relationships among these species are of both practical and theoretical interest. Previous chloroplast DNA (cpDNA) restriction site studies (Wendel and Albert, 1992; DeJoode, 1992) and analyses of nuclear ribosomal DNA (rDNA) sequences (Cronn *et al.*, 1996; Seelanan *et al.*, 1997; Wendel *et al.*, 1995a) have supported monophyly for the subgenus and have resolved six primary clades corresponding to the recognized taxonomic subsections. There has been little agreement, however, on resolution of relationships within and among these six clades. In the present paper, we explore the utility of a nuclear-encoded alcohol dehydrogenase (*Adh*) gene as an independent estimator of the phylogeny of this group of diploid cottons.

Diploid *Gossypium* species have at least seven *Adh* loci (Small, 1999). A previous study has shown that one of these loci (*AdhC*) was useful for examining phylogenetic relationships among the recently diverged tetraploid species of *Gossypium* (Small *et al.*, 1998). A second study employed another locus (*AdhD*) in a study of phylogeny of a group of diploid Australian *Gossypium* species (Seelanan *et al.*, 1999). In the process of exploring the molecular evolution and potential phylogenetic utility of these genes, we chose a different locus, *AdhA*, for phylogenetic analysis of the D-genome diploid cotton species. A previous study of *AdhA* (Small *et al.*, 1999) has shown that this locus is unusual in that two introns normally found in most plant *Adh* genes are missing from *AdhA* (introns 4 and 7; Fig. 1). Initial Southern hybridization analyses indicated that

[1] To whom correspondence and reprint requests should be addressed at present address: Department of Botany, 437 Hesler Biology, The University of Tennessee, Knoxville, TN 37996-1100. Fax: (423) 974-2258. E-mail: rsmall@utk.edu.



1055-7903/00 $35.00

**TABLE 1**

**Plant Materials Used; Voucher Specimens Deposited at ISC**

| Species | Accession No. |
|---|---|
| *Gossypium* L. | |
| Subgenus *Sturtia* (R. Brown) Todaro | |
| *G. robinsonii* F. Mueller | AZ-50 |
| Subgenus *Gossypium* | |
| *G. herbaceum* L. | A1-73 |
| Subgenus *Houzingenia* (Fryxell) Fryxell | |
| Section *Houzingenia* | |
| Subsection *Houzingenia* | |
| *G. thurberi* Todaro | D1-17 |
| *G. thurberi* | D1-8 |
| *G. trilobum* (DC) Skovsted | No accession no. |
| *G. trilobum* | Galau |
| Subsection *Integrifolia* (Todaro) Todaro | |
| *G. davidsonii* Kellogg | 32 |
| *G. klotzschianum* Andersson | D3k-3 |
| *G. klotzschianum* | D3k-47 |
| Subsection *Caducibracteolata* Mauer | |
| *G. armourianum* Kearney | No accession no. |
| *G. armourianum* | D2-1-6 |
| *G. harknessii* Brandegee | No accession no. |
| *G. turneri* Fryxell | D10-3 |
| *G. turneri* | D10-9 |
| Section *Erioxylum* (Rose & Standley) Prokhanov | |
| Subsection *Erioxylum* | |
| *G. aridum* (Rose & Standley) Skovsted | D4-12 |
| *G. aridum* | DRD-118 |
| *G. aridum* | DRD-176 |
| *G. aridum* | DRD-185 |
| *G. schwendimanii* Fryxell & Koch | DRD-149 |
| *G. schwendimanii* | No accession no. |
| *G. schwendimanii* | DRD-133 |
| *G. schwendimanii* | DRD-147 |
| *G. laxum* Phillips | L. Phillips |
| *G. laxum* | DRD-104 |
| *G. laxum* | DRD-101-3 |
| *G. laxum* | DRD-106-2 |
| *G. lobatum* Gentry | No accession no. |
| *G. lobatum* | DRD-140 |
| *G. lobatum* | DRD-161 |
| *G. lobatum* | DRD-145 |
| Subsection *Austroamericana* Fryxell | |
| *G. raimondii* Ulbrich | No accession no. |
| *G. raimondii* | D5-37 |
| Subsection *Selera* (Ulbrich) Fryxell | |
| *G. gossypioides* (Ulbrich) Standley | D6-5 |
| *G. gossypioides* | D6-2 |

*AdhA* occurred in only one copy per genome in two D-genome diploid species (*G. raimondii* and *G. trilobum*) and in the two A-genome diploid species (*G. herbaceum* and *G. arboreum*) (Small *et al.*, 1999, and data not shown). Additionally, these experiments showed that there are two *AdhA* copies in the allotetraploid species *G. hirsutum* and *G. barbadense* (Small *et al.*, 1999), one from each diploid parental donor. Finally, using $F_2$ populations synthesized for comparative genome mapping purposes (Brubaker *et al.*, 1999), we genetically mapped *AdhA* in both progenitor diploid genomes and in allopolyploid cotton. The results showed that *AdhA* exists at homologous chromosomal loci in all three of these genetic maps (Small, 1999), thus bolstering the interpretation that the *AdhA* sequences reported here reflect orthologous rather than paralogous genetic loci.

## MATERIALS AND METHODS

### *Plant Materials*

Accessions of all 13 D-genome species were sampled and, in most cases, more than one accession of a species was sampled (Table 1). In addition, we included sequences from an African–Asian A-genome diploid (*G. herbaceum* L.) and an Australian C-genome diploid (*G. robinsonii*) as outgroups. Genomic DNAs were isolated from leaf tissue either by a modification of the method of Doyle and Doyle (1987) or Paterson *et al.* (1993).

### *PCR Amplification and Sequencing*

*AdhA* gene fragments were obtained by one of two methods: (1) PCR amplification of multiple loci using degenerate *Adh* primers followed by cloning or (2) amplification of a single PCR product using locus-specific primers followed by direct sequencing. In all cases, PCR amplifications were performed in a 50-$\mu$L reaction with 1 unit *Taq* polymerase (Promega), 1× buffer (Promega), 200 $\mu$M each dNTP, 2.0 mM $MgCl_2$, 10 pmol each primer, and 1 $\mu$l template DNA (ca. 10–100 ng). Amplification was accomplished using a program of 30 cycles of denaturation at 94°C for 1 min, annealing at 50°C for 1 min, and extension at 72°C for 2 min, followed by a final 5-min extension at 72°C; all amplifications were performed in MJ Research thermocyclers.

Sequences of the outgroup species *G. robinsonii* and *G. herbaceum*, as well as two ingroup accessions (*G. raimondii* [no accession no.] and *G. gossypioides* [$D_6$-2]) were isolated via the cloning method. The sequences of *G. robinsonii, G. herbaceum,* and *G. raimondii* were previously reported (Small *et al.*, 1999). These sequences were amplified using the primers P1 (CTG C[G/T]G T[G/T]G CAT GGG A[A/G]G CAG GGA AGC C) and P2 (GCA CAG CCA CAC CCC AAC CCT G), which are homologous to regions in exon 2 and exon 9, respectively, of the maize *Adh* genes (Fig. 1). PCR using these primers resulted in the amplification of multiple *Adh* sequences, which differed in size. This heterogeneous pool of PCR products was ligated into the plasmid vector pGEM-T (Promega) according to the manufacturer's instructions. Resulting colonies were screened for inserts by PCR amplification using the primers P1 and P2 as described (Small *et al.*, 1998), and *AdhA*-containing plasmids were identified by the size of the amplicon. Individual *AdhA*-containing plas-

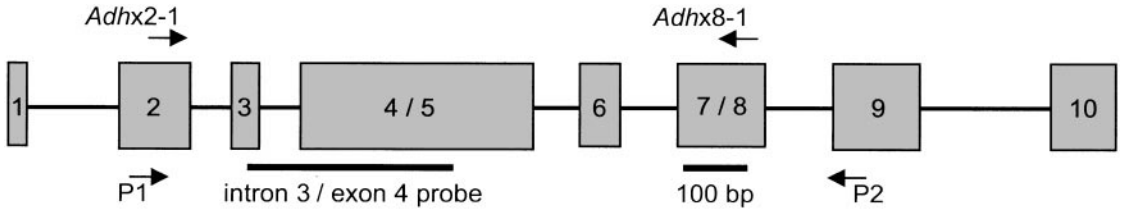


**FIG. 1.** Diagrammatic representation of the *Gossypium AdhA* locus. Exons are shown as shaded boxes and introns as the line connecting the exons. Genomic sequence data are available only for exons 2–8; the lengths of exons 1, 9, and 10 and introns 1 and 9 are extrapolated from other *Gossypium Adh* sequences. PCR amplification primers are shown in their approximate positions (*Adh*x2-1 and *Adh*x8-1 above; P1 and P2 below). The bold line beneath the intron 3/exon 4 region indicates the fragment used as a probe in Southern hybridization analyses. A 100-bp scale bar is shown for reference.

mids were isolated via an alkaline lysis/PEG precipitation miniprep protocol (Sambrook *et al.,* 1989) and were sequenced on an ABI Prism automated DNA sequencer at the Iowa State University DNA Sequencing and Synthesis Facility.

The remaining ingroup sequences (30/32) were isolated using locus-specific PCR primers designed from the above sequences, followed by direct sequencing of the PCR products. The PCR primers *Adh*x2-1 (CTT CAC TGC TTT ATG TCA CAC T) and *Adh*x8-1 (GGA CGC TCC CTG TAC TCC) were used to amplify a ca. 1-kb fragment of the *AdhA* locus. These PCR primers are homologous to regions in exon 2 and exon 8 of the maize *Adh* genes, although the corresponding intron 4 and intron 7 have been deleted in *Gossypium AdhA* (Small *et al.,* 1999; Fig. 1). To eliminate primers and unincorporated dNTPs, PCR products were treated with Shrimp Alkaline Phosphatase and Exonuclease (Amersham) or were purified using Geneclean II (Bio 101). Cleaned PCR products were sequenced directly using the Thermo Sequenase Cycle Sequencing kit (Amersham) and electrophoresed on 5–6% Long Ranger (FMC) polyacrylamide sequencing gels. In cases in which sequences were polymorphic at more than one site, they were deconstructed into individual alleles by cloning the PCR products as described above and sequencing five clones per accession; in all such cases only two sequence types were recovered, indicating that PCR-mediated recombination is unlikely to be an issue. The sequences reported here have been deposited in GenBank under Accession Nos. AF136457–AF136459 and AF182116–AF182153.

### *Southern Hybridizations*

To assess copy number of *AdhA,* we performed Southern hybridization analyses. The number of hybridizing bands can be taken as an estimate of the number of genomic loci for *AdhA* by using locus-specific probes and enzymes that are known to have no recognition sites within the probe region. The presence of more than one band may be indicative of either multiple loci or heterozygosity. By using multiple enzymes, we can differentiate between these two possibilities because allelic variation is expected to be revealed with few enzymes, whereas variation in the number of loci would be expected to be revealed with most enzymes.

Approximately 5 $\mu$g of each genomic DNA was digested separately with *Eco*RI and *Hin*dIII and the resulting fragments were electrophoresed in 1% agarose gels and transferred to nylon membranes (MSI). Membranes were probed with a $^{32}$P-labeled *AdhA* fragment representing the majority of the intron 3 + exon 4 region (Fig. 1). This probe was generated by PCR amplification from a cloned *G. robinsonii AdhA* fragment using primers Fex3 (ATG A[A/G]G C[C/T]G GAG GGT) and Bex4-3′ (CA[A/G] AC[C/T] TT[A/G] TC[A/G] AG) (obtained from B. Gaut, U.C. Irvine) and labeled via random primer labeling (Gibco-BRL). Hybridizations (6× SSC, 0.5% SDS, 100 $\mu$g/ml herring sperm DNA) and washes (0.1× SSC, 0.5% SDS) were performed at 65°C, with intervening rinses in 2× SSC, 0.5% SDS and 2× SSC, 0.1% SDS, both at room temperature, according to Sambrook *et al.* (1989). Selected species were subjected to additional Southern hybridization analysis using the enzymes *Bam*HI, *Eco*RV, and *Xba*I (see below).

### *Analyses*

Sequence alignment was unambiguous and was performed with the assistance of Clustal W (Thompson *et al.,* 1994). Intron/exon boundaries were determined by comparing the genomic sequences obtained to previously published *Adh* exon sequences and noting preservation of the GT . . . AG intron boundaries. In cases in which a sequence from an individual accession was polymorphic at more than one nucleotide position, we resolved the alleles by cloning the PCR products and sequencing multiple clones. This approach allowed us to unambiguously assign linkage relationships among polymorphic nucleotide positions. Insertion/deletion (indel) mutations were scored as binary characters and added to the matrix. Phylogenetic analyses were conducted using both maximum-parsimony and neighbor-joining (Saitou and Nei, 1987) as implemented in PAUP* (Swofford, 1999). Robustness of phylogenetic hypotheses was assessed with the decay index (Bremer,

1988) and the jacknife (implemented in PAUP*; fast-heuristic search with 10,000 replicates, 50% deletion per replicate).

## RESULTS

### *AdhA Sequences*

In this study, *AdhA* sequences were obtained from 32 ingroup accessions, representing the 13 D-genome diploid species of *Gossypium* subgenus *Houzingenia*, and from two outgroup species (*G. robinsonii* and *G. herbaceum*). Typically, plant *Adh* sequences contain 10 exons and 9 introns (e.g., Dennis *et al.*, 1984, 1985; Perry and Furnier, 1996; Sang *et al.*, 1997; Trick *et al.*, 1988), although there are exceptions, such as *Arabidopsis thaliana*, with only 7 exons and 6 introns (Chang and Meyerowitz, 1986). The *Gossypium AdhA* sequences reported here (which span the coding region from exon 2 to exon 8) are lacking introns 4 and 7, which appear to have been precisely removed (Small *et al.*, 1999; Fig. 1). Assuming retention of regions homologous to canonical angiosperm exons 1, 9, and 10, which are external to the amplification primers, *Gossypium AdhA* genes are postulated to have only 8 exons.

The aligned sequence data matrix consisted of 40 terminals (i.e., alleles) and 996 characters, the latter comprising 989 nucleotides (nt) and seven indels. A matrix of variable sites is shown in Fig. 2. Absolute sequence lengths ranged from 982 to 988 nt among the alleles sampled. In the aligned matrix, 662 nt were in exons and 327 nt were in introns. Within the ingroup sequences, there were 46 variable nucleotide positions (4.7%), of which 37 (3.7%) were cladistically informative. Seven indels were introduced during sequence alignment of which six were cladistically informative. All indels occurred within introns; five consisted of 1-nt indels (primarily in runs of A or T) and the remaining two were 3 nt in length. Of the 46 variable ingroup sites, 23 occurred within introns and 23 were in exons. Of the 23 substitutions in exons, 7 were nonsynonymous (amino acid-altering) substitutions, 3 of which were autapomorphic. Thus, there were 4 nonsynonymous substitutions that could potentially be under selective pressures that served as synapomorphies. Reanalysis of the data set with these 4 substitutions excluded resulted in a tree topology congruent with that of the original analyses.

### *Copy Number Analyses*

Initial Southern hybridization analyses using an *AdhA*-specific probe and *Eco*RI- and *Hin*dIII-digested DNAs of members of each taxonomic subsection revealed a potential gene duplication in the species of subsection *Erioxylum* (Fig. 3). Because there are no restriction sites in the probe region in any of our sequences for the enzymes used, we inferred that each band on the autoradiograph corresponds to a single *AdhA* locus. To explore the possibility that the multiple bands observed are due to heterozygosity rather than gene duplication, we performed additional Southern hybridization analyses for the four species of subsection *Erioxylum* using the restriction enzymes *Bam*HI, *Eco*RV, and *Xba*I (Fig. 4). All species of subsection *Erioxylum* displayed two or more hybridizing bands with most enzyme digestions, indicating that *AdhA* loci are duplicated in species of subsection *Erioxylum*.

### *Phylogenetic Analyses*

Parsimony analysis (heuristic search with 100 random addition sequences saving 1000 trees per replicate) of individual alleles resulted in 76,000 most-parsimonious trees (length = 75; consistency index [CI] = 0.87; retention index [RI] = 0.92). The strict consensus of these trees is shown in Fig. 5A. Neighbor-joining analysis employing Jukes–Cantor (1969) corrected distances (gaps treated as missing data) was also performed; the resulting tree is shown in Fig. 5B. These analyses reveal a well-resolved, though weakly supported, phylogenetic hypothesis. Ingroup taxa fall into four primary clades: (1) *G. gossypioides* (=subsection *Selera*); (2) *G. thurberi* and *G. trilobum* (=subsection *Houzingenia*); (3) *G. raimondii* (=subsection *Austroamericana*), *G. turneri, G. armourianum, G. harknessii* (=subsection *Cauducibracteolata*), *G. davidsonii,* and *G. klotzschianum* (=subsection *Integrifolia*); and (4) *G. aridum, G. laxum, G. lobatum,* and *G. schwendimanii* (=subsection *Erioxylum*). As shown by the almost uniformly low decay and jacknife values (Fig. 5A) and short interior branches (Fig. 5B), few character-state changes support most aspects of this inferred history.

For 11 of the 13 ingroup species (all but *G. davidsonii* and *G. harknessii*), at least two accessions were sampled. In some cases, *AdhA* alleles from individual species were more closely related to each other than to alleles of any other taxon (e.g., *G. raimondii, G. trilobum*), whereas in other species allelic monophyly was not directly contradicted by the polytomies formed by alleles from closely related species (e.g., the 3 species from Baja California, *viz., G. turneri, G. harknessii,* and *G. armourianum;* Fig. 5). The notable exception was for species of subsection *Erioxylum*. Although alleles from *G. aridum, G. laxum, G. lobatum,* and *G. schwendimanii* collectively comprise a monophyletic clade, no taxon within this clade has a monophyletic collection of alleles.

## DISCUSSION

### *Phylogeny of D-Genome Gossypium Species*

The phylogenetic hypothesis presented in Fig. 5 is congruent with current taxonomic classification (Fryx-

| Species | Accession | 4 | 3 0 | 3 4 | 3 5 | 4 1 | 4 2 | 4 3 | 4 5 | 5 5 | 5 6 | 5 7 | 6 1 | 6 8 | 7 3 | 7 7 | 8 5 | 8 9 | 9 3 | 9 6 | 1 2 7 | 1 4 5 | 1 6 5 | 1 8 0 | 1 8 1 | 2 0 1 | 2 2 3 | 2 3 1 | 2 4 9 | 2 7 0 | 2 8 0 | 2 8 2 | 2 8 5 | 3 0 3 | 3 3 6 | 3 4 2 | 3 4 3 | 4 4 3 | 4 4 8 | 4 5 3 | 5 0 7 | 5 1 7 | 5 5 3 | 5 7 2 | 5 6 9 | 6 0 7 | 6 6 2 | 6 6 5 | 6 7 3 | 7 0 4 | 7 2 6 | 7 7 8 | 8 3 8 | 8 8 3 | 8 8 6 | 8 8 7 | 8 8 8 | 8 8 9 | 8 9 1 | 9 0 3 | 9 2 5 | 9 6 9 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *G. robinsonii* | AZ-50 | G | A | T | A | T | A | A | G | A | C | A | A | A | T | T | T | T | T | T | A | C | - | T | T | A | T | A | T | G | G | C | T | A | G | T | T | C | T | T | T | C | G | G | T | T | G | T | T | C | T | T | G | T | A | T | G | G | T | T | T | - | T | A | T | A | C | C |
| *G. herbaceum* | A1-73 | . | . | C | . | A | . | . | . | . | . | . | . | . | . | . | . | . | . | . | T | . | A | . | . | . | G | . | . | . | T | . | . | . | . | . | . | . | . | . | . | . | A | C | . | C | . | ? | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | - | . | . | . | . | . | . |
| *G. raimondii* | | . | . | T | . | . | . | . | . | . | . | . | . | G | . | - | C | C | . | C | T | . | - | - | C | T | . | . | C | . | A | . | . | . | . | . | T | . | C | . | G | T | A | . | . | . | . | . | A | . | . | . | . | . | C | . | . | . | . | G | G | C | G | . | . | . | . |
| *G. raimondii* | D5-37 | . | . | T | . | . | . | . | . | . | . | . | . | G | . | - | C | . | C | T | . | - | - | C | T | . | . | . | . | A | . | . | . | . | . | T | . | C | . | G | T | A | . | . | . | . | . | A | . | . | . | . | . | C | . | . | . | . | G | G | C | . | . | . | . | ? |
| *G. gossypioides* | D6-5 | . | . | . | . | C | . | . | . | C | . | . | . | . | . | - | . | C | . | . | T | . | - | . | C | . | . | . | . | A | . | . | . | . | . | . | . | . | . | . | . | . | A | . | . | . | . | . | A | . | . | . | . | . | . | . | . | . | . | G | - | . | . | . | . | . | . |
| *G. gossypioides* | D6-2 | . | . | . | . | C | . | . | . | C | . | . | . | . | . | - | . | . | . | . | T | . | - | - | C | . | . | . | . | A | . | . | . | . | . | . | . | . | . | . | . | . | A | . | . | . | . | . | A | . | . | . | . | . | . | . | . | . | . | G | - | . | . | . | . | . | ? |
| *G. turneri* | D10-3 | . | . | T | . | . | . | . | . | . | . | . | . | . | . | - | . | C | . | . | T | . | - | - | C | T | . | . | . | A | . | . | . | . | . | . | . | C | . | . | . | A | . | . | . | . | . | A | . | . | . | . | . | . | . | . | . | . | G | - | . | . | . | . | . | ? |
| *G. turneri* | D10-9 | . | . | T | . | . | . | . | . | . | . | . | . | . | . | - | . | C | . | . | T | . | - | - | C | T | . | . | . | A | . | . | . | . | . | . | . | C | . | . | . | A | . | . | . | . | . | A | . | . | . | . | . | - | - | - | . | . | G | - | . | . | . | . | . | . |
| *G. armourianum* | | . | . | T | . | . | . | . | . | . | . | . | . | . | . | - | . | C | . | . | T | . | - | - | C | T | . | . | . | A | . | . | . | . | . | . | . | C | . | . | . | A | . | . | . | . | . | A | . | . | . | . | . | . | . | . | . | . | G | - | . | . | . | . | . | ? |
| *G. armourianum* | D2-1-6 | . | . | T | . | . | . | . | . | . | . | . | . | . | . | - | . | C | . | . | T | . | - | - | C | T | . | . | . | A | . | . | . | . | . | . | . | C | . | . | . | A | . | . | . | . | . | A | . | . | . | . | . | . | . | . | . | . | G | - | . | . | . | . | . | . |
| *G. harknessii* | A | . | . | T | . | . | . | . | . | . | . | . | . | . | . | - | . | C | . | . | T | . | - | - | C | T | . | . | . | A | . | . | . | . | . | . | . | C | . | . | . | A | . | . | . | . | . | A | . | . | C | . | . | . | . | . | . | . | G | - | . | . | . | . | . | ? |
| *G. harknessii* | B | . | . | T | . | . | . | . | . | . | . | . | . | . | . | - | . | C | . | . | T | . | - | - | C | T | . | . | . | A | . | . | . | . | . | . | . | C | . | . | . | A | . | . | . | . | . | A | . | . | . | . | . | . | . | . | . | . | G | - | . | . | . | . | . | ? |
| *G. thurberi* | D1 #17 | . | . | T | . | . | . | . | - | - | - | . | . | . | . | - | . | C | . | . | T | . | - | . | C | T | A | . | . | A | . | . | C | . | . | . | . | . | T | . | A | . | . | . | . | . | A | . | . | . | . | . | A | . | . | . | . | G | - | . | . | . | . | . | ? |
| *G. thurberi* | D1 #8 | . | . | T | . | . | . | . | - | - | - | . | . | . | . | - | . | C | . | . | T | . | - | . | C | T | A | . | . | A | . | . | C | . | . | . | . | . | T | . | A | . | . | . | . | . | A | . | . | . | . | . | A | . | . | . | . | G | - | . | . | . | . | . | T |
| *G. trilobum* | | . | . | T | . | . | . | . | - | - | - | . | . | . | . | - | . | C | . | . | T | . | - | . | C | T | A | . | . | A | . | . | C | . | C | . | . | . | T | . | A | . | . | . | . | . | A | . | . | . | . | . | A | . | . | . | . | G | - | . | . | . | . | . | ? |
| *G. trilobum* | Galau | . | . | T | . | . | . | . | - | - | - | . | . | . | . | - | . | C | . | . | T | . | - | . | C | T | A | . | . | A | . | . | C | . | C | . | . | . | T | . | A | . | . | . | . | . | A | . | . | . | . | . | A | . | . | . | . | G | - | . | . | . | . | . | T |
| *G. klotzschianum* | D3k-3 | . | . | T | . | . | . | . | . | . | . | . | . | . | . | - | . | C | . | . | T | . | - | - | C | T | . | . | . | A | . | C | . | . | . | . | . | C | . | . | . | A | . | . | . | . | . | A | . | C | . | . | . | . | . | . | . | . | G | - | . | . | . | . | . | ? |
| *G. klotzschianum* | D3k-47 | . | . | T | . | . | . | . | . | . | . | . | . | . | . | - | . | C | . | . | T | . | - | - | C | T | . | . | . | A | . | C | . | . | . | . | . | C | . | . | . | A | . | . | . | . | . | A | . | C | . | . | . | . | . | . | . | . | G | - | . | . | . | . | . | . |
| *G. davidsonii* | D3d #32 | . | . | T | . | . | . | . | . | . | . | . | . | . | . | - | . | C | . | . | T | . | - | - | C | T | . | . | . | A | . | C | . | . | . | . | . | C | . | . | . | A | . | . | . | . | . | A | . | C | . | . | . | . | . | . | . | . | G | - | . | . | . | . | . | ? |
| *G. aridum* | D4-12 | A | . | T | . | . | . | . | . | . | . | . | . | . | . | - | . | C | . | . | T | . | - | - | C | T | . | . | . | A | . | . | . | . | . | . | . | . | . | . | . | A | . | . | . | G | A | . | . | . | . | . | . | . | . | . | . | G | - | . | . | . | . | . | . |
| *G. aridum* | DRD118 | A | . | T | . | . | . | . | . | . | . | . | - | . | . | - | . | C | . | . | T | . | - | - | C | T | . | . | . | A | . | . | . | . | . | . | . | . | . | . | . | A | . | . | . | G | A | . | . | . | . | . | . | . | . | . | . | G | - | . | C | . | . | . | . |
| *G. aridum* | DRD185 | A | . | T | . | . | . | . | . | . | . | . | - | . | . | - | . | C | . | . | T | . | - | - | C | T | . | . | . | A | . | . | . | . | . | . | . | . | C | . | . | A | . | . | . | G | A | . | . | . | . | . | . | . | . | . | . | G | - | . | . | . | . | . | . |
| *G. aridum* | DRD176 | A | . | T | . | . | . | . | . | . | . | . | - | . | . | - | . | C | . | . | T | . | - | - | C | T | . | . | . | A | . | . | . | . | . | . | . | . | . | . | . | A | . | . | . | G | A | . | . | . | . | . | . | . | . | . | . | G | - | . | . | . | . | . | . |
| *G. schwendimanii* | DRD149 | . | . | T | . | . | . | . | . | . | . | . | . | A | . | - | . | C | . | . | T | . | - | - | C | T | . | . | . | A | . | . | . | . | . | . | . | C | . | . | . | A | . | . | . | G | A | A | . | . | . | G | A | . | . | . | . | G | - | . | . | . | . | . | ? |
| *G. schwendimanii* | A | A | . | T | . | . | . | . | . | . | . | . | - | . | . | - | . | C | . | . | T | . | - | - | C | T | . | . | . | A | . | . | . | . | . | . | . | . | . | . | . | A | . | . | . | G | A | . | . | . | . | . | . | . | . | . | . | G | - | . | . | . | . | . | . |
| *G. schwendimanii* | B | . | C | . | T | . | . | . | . | . | . | . | . | . | . | - | . | C | . | . | T | . | - | - | C | T | . | T | . | A | . | . | . | . | . | . | . | . | . | . | . | A | . | . | . | G | A | . | . | . | . | . | . | . | . | . | . | G | - | . | . | . | . | . | . |
| *G. schwendimanii* | DRD133 | . | . | T | . | . | . | . | . | . | . | . | . | A | . | - | . | C | . | . | T | . | - | - | C | T | . | . | . | A | . | . | . | . | . | C | . | . | . | . | . | A | . | . | . | G | A | A | . | . | . | G | A | . | . | . | . | G | - | . | . | . | . | . | . |
| *G. schwendimanii* | DRD147-A | A | . | T | . | . | . | . | . | . | . | . | - | . | . | - | . | C | . | . | T | . | - | - | C | T | . | . | . | A | . | . | G | . | C | . | . | . | . | . | . | A | . | C | . | G | A | . | . | C | . | . | . | . | . | . | . | G | - | . | C | . | . | . | . |
| *G. schwendimanii* | DRD147-B | . | . | T | . | . | . | . | . | . | . | . | . | A | . | - | . | C | . | . | T | . | - | - | C | T | . | . | . | A | . | . | . | . | C | . | . | C | . | . | . | A | . | . | . | G | A | . | . | . | . | G | A | . | . | . | . | G | - | . | . | . | . | . | . |
| *G. laxum* | DRD104-A | . | . | T | . | . | . | . | . | . | . | . | . | . | . | - | . | C | G | . | T | T | - | . | C | T | . | T | . | A | . | . | . | . | . | . | . | . | . | . | . | A | . | . | . | G | A | . | . | . | . | . | . | . | . | . | . | G | - | . | . | . | . | . | . |
| *G. laxum* | DRD104-B | . | . | T | . | . | A | . | . | . | . | . | . | A | . | - | . | C | . | . | T | T | - | . | C | T | . | T | . | A | . | . | . | . | . | . | . | . | . | . | . | A | . | . | . | G | A | . | . | . | . | . | . | . | . | . | . | G | - | . | . | . | . | . | . |
| *G. laxum* | L. Phillips-A | . | . | T | . | . | T | . | . | . | . | . | . | . | . | - | . | C | . | . | T | . | - | - | C | T | . | T | . | A | . | . | . | . | . | . | . | . | . | . | . | A | . | . | . | G | A | . | . | . | . | . | . | . | . | . | . | G | - | . | . | G | . | . | . |
| *G. laxum* | L. Phillips-B | . | . | T | . | . | . | . | . | . | . | . | . | . | . | - | . | C | . | . | T | . | - | - | C | T | . | T | . | C | A | . | . | . | . | . | . | . | . | . | . | A | . | . | . | G | A | . | . | . | . | . | . | . | . | . | . | G | - | . | . | . | T | . | . |
| *G. laxum* | DRD101-3 | . | . | T | . | . | . | . | . | . | . | . | . | . | . | - | . | C | . | . | T | . | - | - | C | T | . | T | . | C | A | . | . | . | . | . | . | . | . | . | . | A | . | . | . | G | A | . | . | . | . | . | . | . | . | . | . | G | - | . | . | . | . | . | . |
| *G. laxum* | DRD106-2-A | . | . | T | . | . | . | . | . | . | . | . | . | . | . | - | . | C | . | . | T | . | - | - | C | T | . | . | . | C | A | . | . | . | . | C | . | . | . | . | . | A | . | . | . | G | A | . | . | . | . | . | . | . | . | . | . | G | - | . | . | . | T | . | . |
| *G. laxum* | DRD106-2-B | . | . | T | . | . | A | . | . | . | . | . | . | A | . | - | . | C | . | . | T | . | - | - | C | T | . | . | . | A | . | . | . | . | . | C | . | . | . | . | . | A | . | . | . | G | A | . | . | . | . | . | . | . | . | . | . | G | - | . | . | . | . | . | . |
| *G. lobatum* | | A | . | T | . | . | . | . | . | . | . | . | - | . | . | - | . | C | . | . | T | . | - | - | C | T | . | . | . | A | . | . | . | . | . | . | . | . | . | . | . | A | . | . | . | G | A | . | . | . | . | . | . | . | . | . | . | G | - | . | C | . | . | . | . |
| *G. lobatum* | DRD140 | A | . | T | . | . | . | . | . | . | . | . | - | . | . | - | . | C | . | . | T | . | - | - | C | T | . | . | . | A | . | . | . | . | . | . | . | . | . | . | . | A | . | . | . | G | A | . | . | . | . | . | . | . | . | . | . | G | - | . | C | . | . | . | . |
| *G. lobatum* | DRD161 | A | . | T | . | . | . | . | . | . | . | . | - | . | . | - | . | C | . | . | T | . | - | - | C | T | . | . | . | A | . | . | . | . | . | . | . | . | . | . | . | A | . | . | . | G | A | . | . | . | . | . | . | . | . | . | . | G | - | . | C | . | . | . | . |
| *G. lobatum* | DRD145-A | A | . | T | . | . | . | . | . | . | . | . | - | . | . | - | . | C | . | . | T | . | - | - | C | T | . | . | . | A | . | . | . | . | . | . | . | . | C | . | . | A | . | . | . | G | A | . | . | . | A | . | . | . | . | . | . | G | - | . | C | . | . | . | . |
| *G. lobatum* | DRD145-B | A | . | T | . | . | . | . | . | . | . | . | - | . | . | - | . | C | . | . | T | . | - | - | C | T | . | . | . | A | . | . | . | . | . | . | . | . | C | . | . | A | . | . | . | G | A | . | . | . | . | . | . | . | . | . | . | G | - | . | C | . | . | . | . |

**FIG. 2.** Matrix of variable sites in *Gossypium AdhA* sequences. All sequences are shown relative to the *Gossypium robinsonii* sequence; a "." indicates identity with the *G. robinsonii* sequence; a "-" indicates a gap inserted for alignment. Above each nucleotide is its numbered position in the aligned sequence.

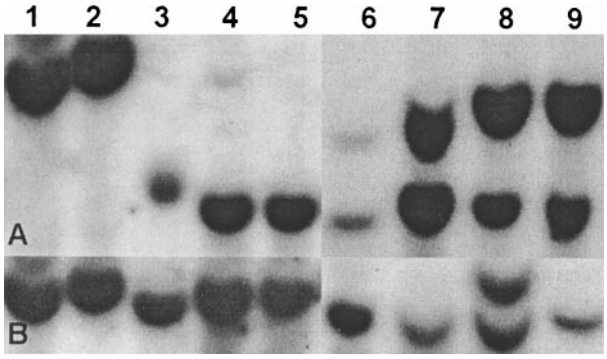


**FIG. 3.** Composite of Southern blot analysis of representatives of the taxonomic subsections of the D-genome *Gossypium* species. Lanes: 1, *G. harknessii;* 2, *G. klotzschianum;* 3, *G. gossypioides;* 4, *G. raimondii;* 5, *G. trilobum;* 6, *G. aridum;* 7, *G. laxum;* 8, *G. lobatum;* 9, *G. schwendimanii.* (A) and (B) show digestions with *Eco*RI and *Hin*dIII, respectively, probed with a *G. robinsonii AdhA* intron 3/exon 4 fragment. In both A and B, lanes 1–5 and 6–9, respectively, are from different hybridization experiments, although other blots (not shown) with all species on the same membrane gave identical results.

ell, 1992; Table 1) in that subsections are monophyletic (e.g., subsection *Houzingenia—G. trilobum* and *G. thurberi*) or their monophyly is not contradicted (e.g., subsection *Caducibracteolata—G. armourianum, G. harknessii,* and *G. turneri*). Sectional monophyly, however, is contradicted for both of the two sections (Fig. 5). Specifically, the components of section *Houzingenia,* namely subsections *Cauducibracteolata, Houzingenia,* and *Integrifolia,* are found dispersed in two of the four primary clades, whereas members of section *Erioxylum,* i.e., subsections *Austroamericana, Erioxylum,* and *Selera,* are found in three of the four primary clades. Not only is there inconsistency between present classification and the *AdhA* phylogeny, but also phylogenetic estimates based on previous cpDNA and rDNA data are incongruent with the existing taxonomic framework as well as with each other (DeJoode, 1992; Wendel and Albert, 1992; Wendel *et al.,* 1995a). To some extent, this is not surprising, given the relatively low support in each of the existing data sets for many of the major lineages resolved. This combination of weak support of contradictory relationships suggests that the radiation of the D-genome cottons into their respective lineages took place relatively rapidly, thus confounding attempts to reconstruct their history.

*Gossypium gossypioides*

While the overall relationships within the D-genome cottons were not well resolved in our analyses, one phylogenetic result deserves highlighting. This concerns the placement of the enigmatic species *G. gossypioides,* which resolves as sister to the remaining 12 D-genome species. Previous analyses of cpDNA restriction site variation placed *G. gossypioides* as the sister taxon to *G. raimondii* (Wendel and Albert, 1992) or as sister to a (*G. raimondii* (*G. aridum, G. laxum, G. lobatum, G. schwendimanii*)) clade (DeJoode, 1992). Analysis of ITS sequences placed *G. gossypioides* in a clade with the African–Asian, A-genome diploids, due to apparent introgression and recombination between the rDNA sequences of *G. gossypioides* and an A-genome species (Wendel *et al.,* 1995a). Analysis of 5S rDNA sequences place *G. gossypioides* at the base of the D-genome clade (Cronn *et al.,* 1996), similar to the results obtained here. A previous suggestion (Wendel *et al.,* 1995a) that the genome of *G. gossypioides* is extensively introgressed and recombined with A-genome nuclear material was recently extended by the demonstration that the nuclear genome of *G. gossypioides* contains a diverse assemblage of repetitive DNAs that are otherwise restricted to A-genome cottons (Zhao *et al.,* 1998).

The basal phylogenetic placement of *G. gossypioides* in *AdhA* and 5S rDNA data sets remains difficult to reconcile with the *G. raimondii*-like plastome in *G. gossypioides,* since the latter species is from Mexico and *G. raimondii* is restricted to Peru, as well as with other data supporting a close relationship with *G. raimondii* (Wendel *et al.,* 1995a). On the other hand, if the placement of *G. gossypioides* as derived and sister to *G. raimondii* (or *G. raimondii* + subsection *Erioxylum*) is correct, then it is similarly difficult to envision how *G. gossypioides* acquired "plesiomorphic" *AdhA* and 5S rDNA sequences. At present, this conundrum remains insoluble, but the suggestion emerges that the species has an especially complex and reticulate history.

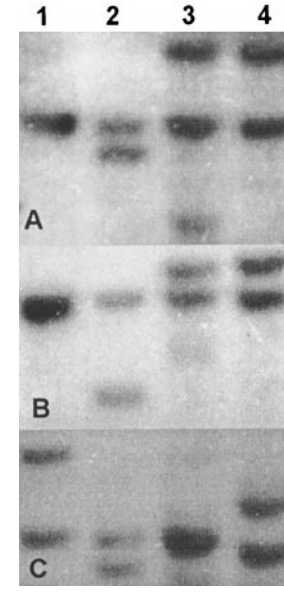


**FIG. 4.** Southern blot analysis of the species of subsection *Erioxylum.* Lanes: 1, *G. aridum;* 2, *G. laxum;* 3, *G. lobatum;* 4, *G. schwendimanii.* (A), (B), and (C) show digestions with *Bam*HI, *Eco*RV, and *Xba*I, respectively, probed with a *G. robinsonii AdhA* intron 3/exon 4 fragment.

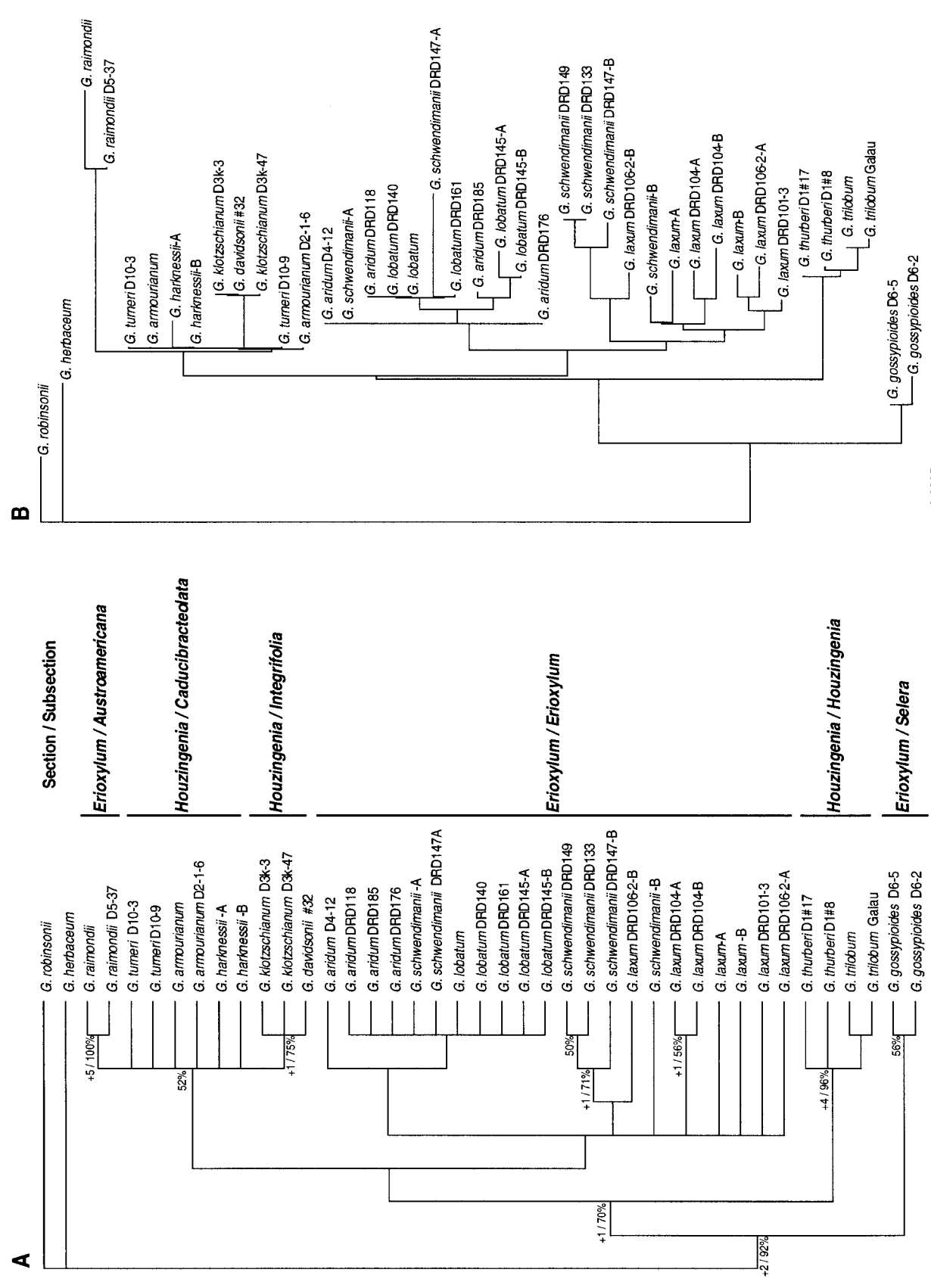


**FIG. 5.** Results of phylogenetic analyses of *AdhA* sequences from D-genome *Gossypium* species and outgroups. Sequences followed by an "A" or "B" indicate deconstructed heterozygosity. (A) The strict consensus of 76,000 equally parsimonious trees with a length of 75, consistency index of 0.87, and retention index of 0.92. Numbers above branches indicate the number of extra steps the node survives (decay value) followed by a jacknife percentage based on 10,000 replicates; only jacknife percentages greater than 50% are shown. (B) A phylogram derived from neighbor-joining analysis of the *AdhA* sequences. The scale bar indicates a Jukes–Cantor distance of 0.0005.

*Structure and Evolution of AdhA in Gossypium*

Alcohol dehydrogenase genes have been characterized from many angiosperms, and with few exceptions, the number and placement of introns is highly conserved (e.g., Dennis *et al.*, 1985, 1986; Perry and Furnier, 1996; Sang *et al.*, 1997; Trick *et al.*, 1988). The exceptions all appear to fall into the category of intron loss. For example, several members of the Brassicaceae have lost *Adh* introns, including *Arabidopsis* (Chang and Meyerowitz, 1986), *Arabis* (Miyashita *et al.*, 1996), and *Leavenworthia* (Charlesworth *et al.*, 1998). In addition, barley has lost intron 9 of *Adh2* (Trick *et al.*, 1988). We add to this list the *AdhA* locus in *Gossypium*, which has lost introns 4 and 7 (Fig. 1). The mechanism(s) responsible for intron loss are not well characterized, although recombination or gene conversion between a cDNA or processed pseudogene and a functional copy is one potential process (Häger *et al.*, 1996; Charlesworth *et al.*, 1998; Drouin and Moniz de Sá, 1997; Loguercio and Wilkins, 1998). The fact that pine *Adh* sequences share the 10 exon/9 intron structure (Perry and Furnier, 1996) with angiosperms suggests that this configuration is ancestral and that the absence of introns in some sequences represents intron loss.

*Intraspecific Variation within Subsection Erioxylum*

One of the significant results of this study was the finding of intraspecific variation among the species of subsection *Erioxylum* (*G. aridum*, *G. laxum*, *G. lobatum*, and *G. schwendimanii*), both in terms of apparent heterozygosity and in allelic diversity. Of the 32 ingroup accessions sampled, seven individuals showed nucleotide polymorphism; six of these were in subsection *Erioxylum*. This result was surprising in that previous results have indicated that genetic diversity in *Gossypium* species is usually low (e.g., Brubaker and Wendel, 1994; Small *et al.*, 1999; Wendel *et al.*, 1989, 1992; Wendel and Percy, 1990). It may be that species in subsection *Erioxylum* possess a distinctive suite of life-history features that promote greater retention of genetic diversity, relative to other wild D-genome diploids. These features may include greater longevity, larger population sizes, and interspecific gene flow, as discussed below.

A related and particularly noteworthy feature of the *AdhA* allelic diversity in subsection *Erioxylum* is the observation that species do not contain monophyletic collections of alleles; i.e., an allele found in one species may be more closely related to an allele from a different species than it is to other alleles from the same species (Fig. 5). Based on morphological evidence, the four species of subsection *Erioxylum* are closely related but also are clearly distinguishable from each other based on a consistent suite of characters (Fryxell, 1988, 1992). Thus, we assume that shared allelic polymorphism does not result from unrealistic taxonomic splitting but instead reflects one or more underlying population genetic and/or molecular evolutionary phenomena. Possible mechanism(s) responsible for the lack of *AdhA* allelic monophyly include gene flow, lineage sorting/noncoalescence, and gene duplication. Each of these possibilities is discussed below.

*Gene flow.* One of the best-documented sources of transspecific sharing of alleles in plants is hybridization and/or introgression (Rieseberg and Wendel, 1993). With respect to the *Gossypium AdhA* data, at least one accession of *G. schwendimanii* (DRD-147) appears to represent a hybrid or an introgressed individual. Specifically, the sampled individual is heterozygous, with one allele apparently contributed by *G. lobatum* and the second allele grouping closely with other *G. schwendimanii* sequences (Fig. 5). This pattern suggests that the *G. schwendimanii* accession DRD-147 contains an *AdhA* sequence introgressed from *G. lobatum*. Notably, this *G. schwendimanii* population was growing sympatrically with *G. lobatum*, and hybrids were apparently present (D. DeJoode, pers. comm.). A second polymorphic accession of *G. schwendimanii* (no accession no.) shows a similar pattern, with one sequence cladistically aligning with *G. aridum* and *G. lobatum* and the other appearing in a clade with *G. laxum* and the remaining *G. schwendimanii* accessions. The origin of this accession is unknown; so, it is not clear whether heterozygosity in this accession reflects hybridization or some other phenomenon. While *Gossypium* species are primarily self-pollinated (Wendel, 1995), outcrossing does occur and interspecific gene flow has been documented in a number of cases (e.g., Brubaker *et al.*, 1993; Wendel *et al.*, 1989; Wendel and Percy, 1990).

*Noncoalescence.* The term lineage sorting is associated with inheritance of polymorphism from a common ancestor in different taxa, followed by fixation of different alleles in the different taxa. Noncoalescence, on the other hand, refers to maintenance of ancestral polymorphism in one or more derived taxa (i.e., alleles coalesce prior to species divergence). This phenomenon can result in alleles that are phylogenetically more closely related to those of other species than to other alleles of the same species. Noncoalescence will be exaggerated in cases of balancing polymorphism or those in which high levels of polymorphism are selectively advantageous, as in self-incompatibility loci (e.g., Klein *et al.*, 1998; Nagl *et al.*, 1998; Richman and Kohn, 1999). Population genetic parameters that determine time of fixation of alleles (e.g., effective population size, selection) will impact both the rate and pattern of molecular evolution. Recent empirical evidence for noncoalescence of alleles includes many examples from maize and other species of *Zea*, including *Adh1* (Gaut and Clegg, 1993), *Adh2* (Goloubinoff *et al.*, 1993), *C1* (Hanson *et al.*, 1996), *glb1* (Hilton and Gaut, 1998), and

ITS (Buckler and Holtsford, 1996a,b). A similar process appears responsible for patterns of *Adh* and *PgiC* allelic relationships in *Leavenworthia* (Charlesworth *et al.*, 1998; Liu *et al.*, 1999) and for 5S ribosomal DNA sequences in *Gossypium* (Cronn *et al.*, 1996).

Inferring noncoalescence requires evidence of one or more ancestral polymorphisms that are shared by two or more species. While this evidence may also be explained by gene flow, as discussed above, we note that the topologies of Fig. 5 do not require gene flow as the causal mechanism for lack of monophyly within species. With respect to species in subsection *Erioxylum*, one might posit that monophyly of *AdhA* alleles from *G. aridum* DRD185, *G. lobatum* DRD145A, and *G. lobatum* DRD145B (Fig. 5B), for example, reflects retention of ancestral alleles into two descendant lineages (the species *G. lobatum* and *G. aridum*) rather than species divergence history. Other, similar examples of noncoalescenceare are readily constructed from Fig. 5, each of which could explain the conflict between the *AdhA* gene tree and the speciation events.

*Gene duplication.* Gene duplication is a potent force in genomic evolution (Ohta, 1994; Ford *et al.*, 1995; Gottlieb and Ford, 1996, 1997). With respect to phylogeny reconstruction, gene duplication creates a requirement for distinguishing orthologous from paralogous genes (Ford *et al.*, 1995; Gottlieb and Ford, 1996, 1997; Sang *et al.*, 1997). It is widely recognized that inadvertent inclusion of paralogous sequences in a phylogenetic analysis could result in a well-supported but misleading topology (Wendel and Doyle, 1998). The situation becomes more complex if the duplicated genes are subject to interlocus concerted evolution (Sanderson and Doyle, 1992). If concerted evolution is strong, then duplicated genes will be homogenized to a single type; if concerted evolution is absent, then duplicated genes will maintain their identity and construction of an orthology–paralogy (OP) tree is possible (Wendel and Doyle, 1998). If, however, concerted evolution is present but weak, then some copies (or parts) of duplicated genes may be homogenized, whereas others are maintained as distinct. Accurate reconstruction of a species tree from such genes is fraught with difficulties (Sanderson and Doyle, 1992; Cronn *et al.*, 1996).

Southern hybridization analyses of D-genome *Gossypium* species revealed an apparent gene duplication in the species of subsection *Erioxylum*. Using a locus-specific probe, most species exhibited a single hybridizing band in *Eco*RI and *Hin*dIII digests (Fig. 3), suggesting that they contain a single genetic locus. Members of subsection *Erioxylum* (*G. aridum*, *G. laxum*, *G. lobatum*, and *G. schwendimanii*), however, exhibited multiple hybridizing fragments suggestive of a gene duplication (Figs. 3 and 4). While the allelic polymorphism noted above for species in subsection *Erioxylum* is not in itself evidence of gene duplication, gene duplication, as evidenced by the Southern hybridization data, could account for the observed elevated polymorphism relative to other D-genome species.

One way to reconcile the Southern hybridization and sequencing data is to posit that a single gene duplication event occurred prior to the diversification of the species of subsection *Erioxylum* but after the separation of this lineage from other D-genome diploids. Under this scenario, each of the extant species of subsection *Erioxylum* would have inherited both *AdhA* copies. One might suppose that if each species inherited two *AdhA* copies, then phylogenetic analysis should reveal a full orthology–paralogy tree (Sanderson and Doyle, 1992), whereby two clades are recovered representing the two genes, within which similar relationships among species are depicted (Ford *et al.*, 1995; Gottlieb and Ford, 1996; Small *et al.*, 1998; Wendel and Doyle, 1998). This expectation is not realized in the present analysis (Fig. 5). In fact, neighbor-joining analysis separates accessions into two weakly supported clades, one consisting primarily of *G. aridum* and *G. lobatum* and the other primarily of *G. schwendimanii* and *G. laxum*.

How are the Southern hybridization data, indicating a gene duplication, and the lack of an OP tree within the subsection to be reconciled? One possibility is that concerted evolution has acted to differentially homogenize paralogous sequences, such that duplicated sequences in *G. schwendimanii* and *G. laxum* have been homogenized to one type and sequences in *G. aridum* and *G. lobatum* have been homogenized to another type (e.g., Wendel *et al.*, 1995b). The two sequences that violate this hypothesis (*G. schwendimanii* DRD-147A and *G. schwendimanii*-A [no accession no.]) could be explained by gene flow, as described above. One problem with this hypothesis is that if concerted evolution were acting to homogenize sequences, we would expect little to no intraspecific polymorphism. This expectation is inconsistent with our results, which show that intraspecific divergence is higher in subsection *Erioxylum* species than in all other sampled species.

A second proposal might be that independent gene duplication events occurred in the *G. schwendimanii/G. laxum* and *G. aridum/G. lobatum* lineages. Relationships among sequences within these groups, however, are not consistent with such a scenario. Specifically, this hypothesis leads to the expectation of two independent OP trees, one within each of these two primary clades. As shown by the topologies recovered (Fig. 5), relationships among sequences do not fit this expectation.

Finally, an additional layer of complexity would be added if recombination, either biological or PCR-mediated, occurred between duplicated loci. As noted above, complete concerted evolution between duplicate copies could account for the lack of an observed OP tree but would also result in low intraspecific diversity, con-

trary to our observations. Limited recombination or gene conversion, however, could potentially result in the disruption of OP relationships and yet maintain observed levels of diversity. Our results suggest that PCR-mediated recombination did not occur. For all samples that displayed polymorphism, we sequenced five clones and obtained only two sequence types which were additive for the initially observed polymorphism. Formal tests of biological recombination are uninformative due to the small number of substitutions that distinguish the sequences.

Available data do not permit a confident resolution of the apparent contradiction between the sequence phylogeny and the Southern hybridization results. One difficulty is that absolute sequence divergence amounts are low, presumably due to the relative recency of the gene duplication event. This confounds efforts to assign alleles to particular paralogues and hence reduces our power to infer the underlying mechanism. In addition, we cannot be certain that our sampling has been complete with respect to alleles and loci. It may be that the allelic phylogeny obtained reflects a combination of processes, including some degree of concerted evolution or recombination, gene flow, and noncoalescence (as discussed above), as well as spurious effects due to low levels of sequence divergence and potentially incomplete sampling. At present, however, we cannot assess the relative importance of these phenomena in generating the *AdhA* gene tree.

One general implication of our results is that gene duplication events may have important yet undetected effects on phylogenetic analyses. We discovered the gene duplication only after sequencing *AdhA* from multiple individuals of the species of subsection *Erioxylum* and finding that the sequences did not form monophyletic groups. As part of our exploration of this phenomenon, we performed Southern hybridization experiments, despite the fact that our initial Southern analyses showed a single copy of *AdhA* per diploid genome in all species sampled. Our results thus add to the growing body of empirical studies in which relatively recent gene duplication events have been detected for plant nuclear genes (e.g., *Clarkia PgiC*—Ford *et al.*, 1995; Gottlieb and Ford, 1996, 1997; *Paeonia Adh*—Sang *et al.*, 1997; *Populus PHYB*—Howe *et al.*, 1998). It may be prudent, therefore, to routinely include Southern hybridization analyses in phylogenetic studies when using nuclear genes.

## CONCLUSIONS

The data presented here propose a well-resolved, but weakly supported phylogenetic hypothesis of relationships among New World diploid *Gossypium* species. While largely congruent with current taxonomy and previous analyses regarding the major clades within this group, there continues to be little agreement across data sets about relationships among taxonomic subsections. We suggest that this results from a relatively recent and rapid radiation of the D-genome lineage into its constituent species-groups. Among the surprising phylogenetic results was the placement of *G. gossypioides* as sister to the remaining D-genome species, a position conflicting with previous biosystematic and cpDNA studies (reviewed in Wendel *et al.*, 1995a) but also found in analysis of 5S rDNA sequences (Cronn *et al.*, 1996). Thus, *AdhA* gene sequence data once again highlight the enigmatic nature of *G. gossypioides* (cf. Wendel *et al.*, 1995a; Zhao *et al.*, 1998) and indicate that this taxon warrants further experimental attention.

The primary molecular evolutionary conclusion is that some D-genome species harbor polyphyletic collections of *AdhA* alleles. We suggest that there are at least three possible explanations for this observation, namely, gene flow, noncoalescence, and gene duplication. Evidence of gene flow is presented for at least one case, as is evidence for a recent gene duplication. Populations of species within subsection *Erioxylum* may also have experienced lineage sorting from polymorphic ancestral populations. Thus, a combination of factors is likely responsible for the lack of allelic monophyly for the four species within the subsection, but at present we cannot assess the relative importance of these (and other?) factors.

Regardless of causal mechanism, shared allelic polymorphisms have implications for phylogenetic analyses (Wendel and Doyle, 1998). Some of the responsible mechanisms, like interspecific hybridization and gene flow, may be expected or at least not unexpected in plants (Rieseberg and Wendel, 1993). Others, however, such as lineage-specific gene duplication, concerted evolution, or long-term noncoalescence, may be unpredictable, given our current level of understanding of molecular evolution of plant nuclear genes (Clegg *et al.*, 1997).

## ACKNOWLEDGMENTS

We are grateful to Dan DeJoode for the initial field and laboratory work that made this project possible and to Ed Percival for providing seed of some species. Research support was provided by the National Science Foundation (to J.F.W.).

## REFERENCES

Bremer, K. (1988). The limits of amino acid sequence data in angiosperm phylogenetic reconstruction. *Evolution* **42:** 795–803.

Brubaker, C. L., Koontz, J. A., and Wendel, J. F. (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., Paterson, A. H., and Wendel, J. F. (1999). Comparative genetic mapping of allotetraploid cotton and its diploid progenitors. *Genome* **42:** 184–203.

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

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

Buckler, E. S., and Holtsford, T. P. (1996b). *Zea* ribosomal repeat evolution and substitution patterns. *Mol. Biol. Evol.* **13:** 623–632.

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

Charlesworth, D., Liu, F.-L., and Zhang, L. (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., Cummings, M. P., and Durbin, M. L. (1997). The evolution of plant nuclear genes. *Proc. Natl. Acad. Sci. USA* **94:** 7791–7798.

Cronn, R. C., Zhao, X., Paterson, A. H., and Wendel, J. F. (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.

DeJoode, D. R. (1992). "Molecular Insights into Speciation in the Genus *Gossypium* L. (Malvaceae)," M.S. thesis, Iowa State Univ., Ames.

Dennis, E. S., Gerlach, W. L., Pryor, A. J., Bennetsen, J. L., Inglis, A., Llewellyn, D., Sachs, M. M., Ferl, R. J., and Peacock, W. J. (1984). Molecular analysis of the alcohol dehydrogenase (*Adh1*) gene of maize. *Nucleic Acids Res.* **12:** 3983–4000.

Dennis, E. S., Sachs, M. M., Gerlach, W. L., Finnegan, E. J., and Peacock, W. J. (1985). Molecular analysis of the alcohol dehydrogenase 2 (*Adh2*) gene of maize. *Nucleic Acids Res.* **13:** 727–743.

Doyle, J. J., and Doyle, J. L. (1987). A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem. Bull.* **19:** 11–15.

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

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

Ford, V. S., Thomas, B. R., and Gottlieb, L. D. (1995). The same duplication accounts for the *PgiC* genes in *Clarkia xantiana* and *C. lewisii* (Onagraceae). *Syst. Bot.* **20:** 147–160.

Fryxell, P. A. (1988). Malvaceae of Mexico. *Syst. Bot. Monographs* **25:** 1–522.

Fryxell, P. A. (1992). A revised taxonomic interpretation of *Gossypium* L. (Malvaceae). *Rheedea* **2:** 108–165.

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

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

Gottlieb, L. D., and Ford, V. S. (1996). Phylogenetic relationships among the sections of *Clarkia* (Onagraceae) inferred from the nucleotide sequences of *PgiC*. *Syst. Bot.* **21:** 45–62.

Gottlieb, L. D., and Ford, V. S. (1997). A recently silenced, duplicate *PgiC* locus in *Clarkia*. *Mol. Biol. Evol.* **14:** 125–132.

Häger, K.-P., Müller, B., Wind, C., Erbach, S., and Fischer, H. (1996). Evolution of legumin genes: Loss of an ancestral intron at the beginning of angiosperm diversification. *FEBS Lett.* **387:** 94–98.

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

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

Howe, G. T., Bucciaglia, P. A., Hackett, W. P., Furnier, G. R., Cordonnier-Pratt, M.-M., and Gardner, G. (1998). Evidence that the phytochrome gene family in Black Cottonwood has one *PHYA* locus and two *PHYB* loci but lacks members of the *PHYC/F* and *PHYE* subfamilies. *Mol. Biol. Evol.* **15:** 160–175.

Jukes, T. H., and Cantor, C. R. (1969). Evolution of protein molecules. *In* "Mammalian Protein Metabolism" (H. N. Munro, Ed.), pp. 21–132. Academic Press, New York.

Klein, J., Sato, A., Nagl, S., and O'huigin, C. (1998). Molecular trans-species polymorphism. *Annu. Rev. Ecol. Syst.* **29:** 1–21.

Liu, F., Charlesworth, D., and Kreitman, M. (1999). The effect of mating system differences on nucleotide diversity at the phosphoglucose isomerase locus in the plant genus *Leavenworthia*. *Genetics* **151:** 343–357.

Loguercio, L. L., and Wilkins, T. A. (1998). Structural analysis of a hmg-coA-reductase pseudogene: Insights into evolutionary processes affecting the *hmgr* gene family in allotetraploid cotton (*Gossypium hirsutum* L.). *Curr. Genet.* **34:** 241–249.

Miyashita, N. T., Innan, H., and Terauchi, R. (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.

Nagl, S., Tichy, H., Mayer, W. E., Takahata, N., and Klein, J. (1998). Persistence of neutral polymorphisms in Lake Victoria cichlid fish. *Proc. Natl. Acad. Sci. USA* **95:** 14238–14243.

Ohta, T. (1994). Further examples of evolution by gene duplication revealed through DNA sequence comparisons. *Genetics* **138:** 1331–1337.

Paterson, A. H., Brubaker, C. L., and Wendel, J. F. (1993). A rapid method for extraction of cotton (*Gossypium* spp.) genomic DNA suitable for RFLP or PCR analysis. *Pl. Mol. Biol. Rep.* **11:** 122–127.

Perry, D. J., and Furnier, G. R. (1996). *Pinus banksiana* has at least seven expressed alcohol dehydrogenase genes in two linked groups. *Proc. Natl. Acad. Sci. USA* **93:** 13020–13023.

Richman, A. D., and Kohn, J. R. (1999). Self-incompatibility alleles from *Physalis:* Implications for historical inference from balanced genetic polymorphisms. *Proc. Natl. Acad. Sci. USA* **96:** 168–172.

Rieseberg, L. H., and Wendel, J. F. (1993). Introgression and its consequences. *In* "Hybrid Zones and the Evolutionary Process" (R. Harrison, Ed.), pp. 70–109. Oxford Univ. Press, New York.

Saitou, N., and Nei, M. (1987). The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4:** 406–425.

Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989). "Molecular Cloning, A Laboratory Manual," 2nd ed, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Sanderson, M. J., and Doyle, J. J. (1992). Reconstruction of organismal and gene phylogenies from data on multigene families: Concerted evolution, homoplasy, and confidence. *Syst. Biol.* **41:** 4–17.

Sang, T., Donoghue, M. J., and Zhang, D. (1997). Evolution of the alcohol dehydrogenase genes in Peonies (*Paeonia*): Phylogenetic relationships of putative nonhybrid species. *Mol. Biol. Evol.* **14:** 994–1007.

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

Seelanan, T., Brubaker, C. L., Stewart, J. McD., Craven, L. A., and Wendel, J. F. (1999). Molecular systematics of Australian *Gossypium* section *Grandicalyx* (Malvaceae). *Syst. Bot.* **24:** 183–208.

Small, R. L. (1999). "Evolution of the Alcohol Dehydrogenase Gene

Family in Diploid and Tetraploid *Gossypium* L," Ph.D. dissertation, Iowa State Univ., Ames.

Small, R. L., Ryburn, J. A., Cronn, R. C., Seelanan, T., and Wendel, J. F. (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.

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

Stewart, J. McD. (1995). Potential for crop improvement with exotic germplasm and genetic engineering. *In* "Challenging the Future: Proceedings of the World Cotton Research Conference-1" (G. A. Constable and N. W. Forrester, Eds.), pp. 313–327. CSIRO, Melbourne.

Swofford, D. L. (1999). PAUP*. Phylogenetic Analysis Using Parsimony (* and Other Methods). Version 4.0b2. Sinauer, Sunderland, MA.

Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994). CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22:** 4673–4680.

Trick, M., Dennis, E. S., Edwards, K. J. R., and Peacock, W. J. (1988). Molecular analysis of the alcohol dehydrogenase gene family of barley. *Pl. Mol. Biol.* **11:** 147–160.

Wendel, J. F. (1995). Cotton. *In* "Evolution of Crop Plants" (J. Smartt and N. W. Simmonds, Eds.), 2nd ed., pp. 358–366. Longman Scientific, Essex, England.

Wendel, J. F., Brubaker, C. L., and Percival, A. E. (1992). Genetic diversity in *Gossypium hirsutum* and the origin of upland cotton. *Am. J. Bot.* **79:** 1291–1310.

Wendel, J. F., and Doyle, J. J. (1998). Phylogenetic incongruence: Window into genome history and molecular evolution. *In* "Molecular Systematics of Plants" (D. Soltis, P. Soltis, and J. Doyle, Eds.), 2nd ed. Chapman & Hall, New York.

Wendel, J. F., Olson, P. D., and Stewart, J. McD. (1989). Genetic diversity, introgression, and independent domestication of Old World cultivated cottons. *Am. J. Bot.* **76:** 1795–1806.

Wendel, J. F., and Percy, R. G. (1990). Allozyme diversity and introgression in the Galapagos Islands endemic *Gossypium darwinii* and its relationship to continental *G. barbadense*. *Biochem. Syst. Ecol.* **18:** 517–528.

Wendel, J. F., and Albert, V. A. (1992). Phylogenetics of the cotton genus (*Gossypium*): 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., Schnabel, A., and Seelanan, T. (1995a). An unusual ribosomal DNA sequence from *Gossypium gossypioides* reveals ancient, cryptic, intergenomic introgression. *Mol. Phylogenet. Evol.* **4:** 298–313.

Wendel, J. F., Schnabel, A., and Seelanan, T. (1995b). Bidirectional interlocus concerted evolution following allopolyploid speciation in cotton (*Gossypium*). *Proc. Natl. Acad. Sci. USA* **92:** 280–284.

Zhao, X.-P., Si, Y., Hanson, R. E., Crane, C. F., Price, H. J., Stelly, D. M., Wendel, J. F., and Paterson, A. H. (1998). Dispersed repetitive DNA has spread to new genomes since polyploid formation in cotton. *Genome Res.* **8:** 479–492.