The tortoise and the hare: choosing between noncoding plastome and nuclear *Adh* sequences for phylogeny reconstruction in a recently diverged plant group[1]

Randall L. Small, Julie A. Ryburn, Richard C. Cronn, Tosak Seelanan, and Jonathan F. Wendel[2]

Department of Botany, Bessey Hall, Iowa State University, Ames, Iowa 50011

Phylogenetic resolution is often low within groups of recently diverged taxa due to a paucity of phylogenetically informative characters. We tested the relative utility of seven noncoding cpDNA regions and a pair of homoeologous nuclear genes for resolving recent divergences, using tetraploid cottons (*Gossypium*) as a model system. The five tetraploid species of *Gossypium* are a monophyletic assemblage derived from an allopolyploidization event that probably occurred within the last 0.5–2 million years. Previous analysis of cpDNA restriction site data provided only partial resolution within this clade despite a large number of enzymes employed. We sequenced three cpDNA introns (*rpl16*, *rpoC1*, *ndhA*) and four cpDNA spacers (*accD-psaI*, *trnL-trnF*, *trnT-trnL*, *atpB-rbcL*) for a total of over 7 kb of sequence per taxon, yet obtained only four informative nucleotide substitutions (0.05%) resulting in incomplete phylogenetic resolution. In addition, we sequenced a 1.65-kb region of a homoeologous pair of nuclear-encoded alcohol dehydrogenase (*Adh*) genes. In contrast with the cpDNA sequence data, the *Adh* homoeologues yielded 25 informative characters (0.76%) and provided a robust and completely resolved topology that is concordant with previous cladistic and phenetic analyses. The enhanced resolution obtained using the nuclear genes reflects an approximately three- to sixfold increase in nucleotide substitution rate relative to the plastome spacers and introns.

Key words: alcohol dehydrogenase; *Gossypium*; molecular phylogenetics; noncoding chloroplast DNA; polyploidy.

The ease of generating DNA sequence data has led to an explosion of molecular phylogenetic analyses in recent years (reviewed in Soltis, Soltis, and Doyle, 1998). In plants, analyses of cpDNA have predominated (reviewed by Olmstead and Palmer, 1994), typically involving genes such as *rbcL*, *matK*, or *ndhF* (e.g., Chase et al., 1993; Olmstead and Palmer, 1994; Olmstead and Sweere, 1994; Steele and Vilgalys, 1994). More recently, sequencing of cpDNA noncoding regions (introns and intergenic spacers) has become popular for analyses at various taxonomic levels (e.g., Morton and Clegg, 1993; Gielly and Taberlet, 1994a, b, 1996; van Ham et al., 1994; Kita, Ueda, and Kadota, 1995; Manen and Natali, 1995; Downie, Katz-Downie, and Cho, 1996; Gielly et al., 1996; Johnson and Hattori, 1996; Jordan, Courtney, and Neigel, 1996; Kelchner and Wendel, 1996; Kelchner and Clark, 1997; Savolainen, Spichiger, and Manen, 1997; Sang, Crawford, and Stuessy, 1997). Noncoding regions have been presumed to be more useful at lower taxonomic ranks because they are less functionally constrained and are therefore freer to vary, thereby potentially providing more phylogenetically informative characters per unit of sequencing effort (Clegg et al., 1994).

One of the often-cited advantages of molecular data for phylogenetic reconstruction is the almost infinite number of characters that can be sampled. Yet, for plant groups where radiations have been relatively recent it may be extraordinarily difficult to generate sufficient phylogenetic signal because of the relatively slow accumulation of mutations, even in “rapidly evolving” noncoding DNA. The literature is replete with cladograms derived from molecular data that are well resolved internally, but that contain unresolved terminal clades of presumably closely related species (e.g., Hodges and Arnold, 1994; Bayer, Hufford, and Soltis, 1996; Soltis et al., 1996; Panero and Jansen, 1997; Sang, Crawford, and Stuessy, 1997). This phenomenon is the focus of the present paper. Specifically, we wished to address the issue of phylogenetic resolution within recent radiations by asking the following questions: (1) are mutation rates sufficiently high in noncoding cpDNA to provide phylogenetic resolution within a group of woody perennials that may be only 0.5–2 million years old? (2) do mutation rates vary among cpDNA noncoding regions, and if so, which exhibits the highest mutation rate? (3) can strictly orthologous low-copy nuclear-encoded genes be isolated, and if so, do they exhibit a higher mutation rate than noncoding cpDNA? (4) what are the relative strengths and weaknesses of the various types of molecular data for evaluating the phylogenetic relationships of recently radiated groups? As a model system for examining these questions we chose the tetraploid species of *Gossypium* L.

Gossypium includes ~50 species (Fryxell, 1992; Wendel, 1995; Wendel, Brubaker, and Seelanan, in press), of which the majority are diploid ($2n = 2x = 26$) and five are allotetraploids ($2n = 4x = 52$). Previous studies have resulted in the phylogenetic hypothesis shown in Fig. 1. The allotetraploid species appear to be a monophyletic

[1] Manuscript received 28 August 1997; revision accepted 13 January 1998.

The authors thank S. Downie, B. Gaut, K. Schierenbeck, T. Barkman, B. Simpson, and R. Wallace for providing PCR primers; R. Jansen and an anonymous reviewer for comments on an earlier draft of the manuscript; and the National Science Foundation for financial support (to JFW).

[2] Author for correspondence (jfw@iastate.edu).

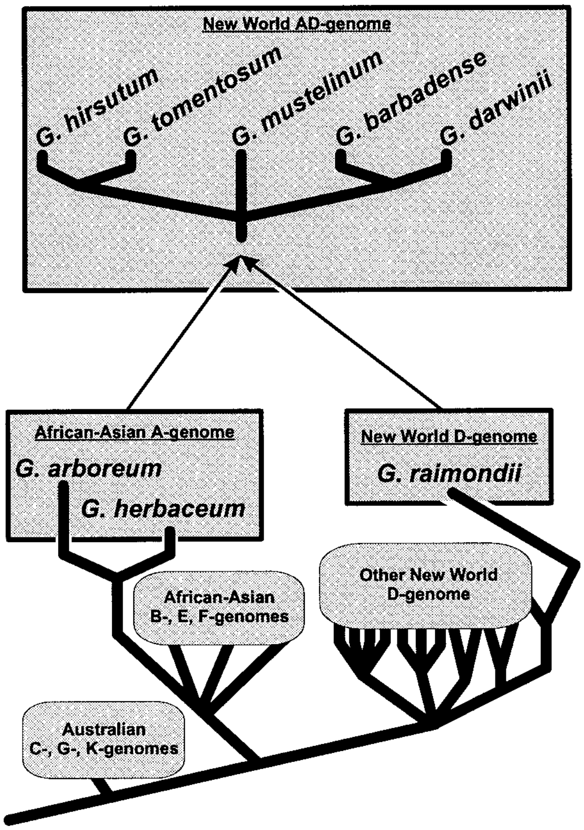

Fig. 1. Phylogenetic hypothesis for intrageneric relationships in *Gossypium*, including the origin of the allotetraploid species. The maternal diploid parent is represented by the extant A-genome species, *G. arboreum* and *G. herbaceum*, while the paternal diploid parent is represented by the extant D-genome species, *G. raimondii*.

assemblage derived from a single polyploidization event ~0.5–2 million years ago (Wendel, 1989; Wendel and Albert, 1992; Seelanan, Schnabel, and Wendel, 1997), and despite extensive efforts directed at understanding relationships among tetraploid cottons, only weak resolution has been obtained (Endrizzi, Turcotte, and Kohel, 1985; Wendel, 1989; DeJoode and Wendel, 1992; Wendel and Albert, 1992; Reinisch et al., 1994; Cronn et al., 1996; Wendel, Schnabel, and Seelanan, 1995a, b; Seelanan, Schnabel, and Wendel, 1997). In addition to cpDNA and rDNA restriction site data, sequences from the nuclear ribosomal ITS regions are available for all tetraploid species (Wendel, Schnabel, and Seelanan, 1995a, b; Seelanan, Schnabel, and Wendel, 1997) and *ndhF* data are available for two of the five species (Seelanan, Schnabel, and Wendel, 1997). Given voluminous data yet little phylogenetic resolution, tetraploid *Gossypium* provides a test case for evaluating the utility of a variety of putatively quickly evolving molecular sequences for resolving the phylogeny of a recent radiation. To this end we sequenced seven cpDNA noncoding regions in each of the five tetraploid species and a representative of the diploid maternal (chloroplast donor; Wendel, 1989) lineage, *G. arboreum* L. In addition, we isolated and sequenced a region of a pair of homoeologous nuclear-encoded alcohol dehydrogenase (*Adh*) genes for these same taxa, as well as a representative of the paternal lineage, *G. raimondii* Ulbrich, and an additional outgroup, *G. robinsonii* F. Mueller.

Table 1. Plant materials. All voucher specimens are deposited at ISC. Voucher abbreviations are as follows: TS = Tosak Seelanan, JFW & TDC = J. F. Wendel and T. D. Couch.

Taxon	Accession	Voucher
C-genome diploid		
Gossypium robinsonii F. Mueller	AZ-50	TS 12
D-genome diploid		
Gossypium raimondii Ulbrich	#436	JFW & TDC 127
A-genome diploid		
Gossypium arboreum L.	A_2-74	JFW & TDC 312
AD-genome tetraploids		
Gossypium hirsutum L.	"Palmeri"	JFW & TDC 632
Gossypium barbadense L.	K101	JFW & TDC 612
Gossypium tomentosum Nuttall ex Seemann	WT936	JFW & TDC 621
Gossypium mustelinum Miers ex Watt	W400	JFW & TDC 622
Gossypium darwinii Watt	WB1215	JFW & TDC 620

MATERIALS AND METHODS

Plant materials and DNA isolation—The species of *Gossypium* studied include one accession from each of the five allotetraploid species and one species from each of three diploid "genome groups." Two of these ("A" and "D" diploids) represent the lineages (maternal and paternal, respectively; Wendel, 1989) from which the allotetraploids were derived, and the third, more distantly related diploid ("C" genome) was included as an outgroup (Table 1). Previous studies support the intrageneric phylogeny shown in Fig. 1 (Wendel and Albert, 1992; Wendel, Schnabel, and Seelanan, 1995a, b; Seelanan, Schnabel, and Wendel, 1997). DNA extractions were carried out as previously described (Paterson, Brubaker, and Wendel, 1993). All sequences obtained in this study have been deposited in GenBank (see GenBank accession numbers given in Table 2).

cpDNA regions—Many cpDNA noncoding regions (introns and intergenic spacers) have been characterized either by direct sequencing (e.g., Morton and Clegg, 1993; van Ham et al., 1994; Manen and Natali, 1995; Downie, Katz-Downie, and Cho, 1996; Gielly et al., 1996; Johnson and Hattori, 1996; Jordan, Courtney, and Neigel, 1996; Kelchner and Wendel, 1996; Kelchner and Clark, 1997; Savolainen, Spichiger, and Manen, 1997; Sang, Crawford, and Stuessy, 1997) or by restriction site analysis of polymerase chain reaction (PCR)-amplified products (Liston, 1992; Rieseberg, Hanson, and Philbrick, 1992; Demesure, Comps, and Petit, 1996; Wolf, Murray, and Sipes, 1997; Wolfe et al., 1997). The regions we chose to study (Table 2, Fig. 2) included both cpDNA introns and intergenic spacers and were selected based on the availability of PCR primers, and/or their size and previous reports of phylogenetic utility. These cpDNA regions all reside in the large single-copy region of the tobacco plastome (Shinozaki et al., 1986) with the exception of the *ndhA* intron, which is in the small single-copy region (Fig. 2). Phylogenetic analyses of sequence data for the cpDNA regions analyzed in this study have been previously reported from other plant groups with the exceptions of the *accD-psaI* spacer and the *ndhA* intron.

The *atpB-rbcL* spacer has been used extensively in phylogenetic and molecular evolutionary analyses (Golenberg et al., 1993; Hodges and Arnold, 1994; Manen, Savolainen, and Simon, 1994; Savolainen et al., 1994; Manen and Natali, 1995; Natali, Manen, and Ehrendorfer, 1995;

Table 2. Regions studied, PCR primer sequences, and GenBank accession numbers.

Region	Primer sequences (written 5′ to 3′)	References	GenBank accession numbers
atpB-rbcL spacer	*atpB*: GTG GAA ACC CCG GGA CGA GAA GTA GT	Hodges and Arnold, 1994	AF031445–AF031450
	rbcL: ACT TGC TTT AGT TTC TGT TTG TGG TGA		
trnL-trnF spacer	E: GGT TCA AGT CCC TCT ATC CC	Taberlet et al., 1991	AF031439–AF031444
	F: ATT TGA ACT GGT GAC ACG AG		
trnT-trnL spacer	A: CAT TAC AAA TGC GAT GCT CT	Tablerlet et al., 1991	AF031433–AF031438
	B: TCT ACC GAT TTC GCC ATA TC		
	trnT-I: CTG ACT CCA TTT TTA TTT TC	this paper	
accD-psaI spacer	*accD*-769F: GGA AGT TTG AGC TTT ATG CAA ATG G	T. Barkman and B. Simpson, University of Texas, Austin, personal communication	AF031580–AF031585
	psaI-75R: AGA AGC CAT TGC AAT TGC CGG AAA		
	*accD*I: GGG CTT TGA CTT TGT GAC		
ndhA intron	*ndhA*-F: GGW CTT CTY ATG KCR GGA TAT RGM TC	this paper	AF031574–AF031579
	ndhA-R: CTG YGC TTC MAC TAT ATC AAC TGT AC		
	ndhA-I: ATT CTG CTT TCG GAT CTG		
rp116 intron	F71: GCT ATG CTT AGT GTG TGA CTC GTT G	Jordan, Courtney, and Neigel, 1996; Kelchner and Wendel, 1996	AF031451–AF031456
	R1661: CGT ACC CAT ATT TTT CCA CCA CGA C		
	R1516: CCC TTC ATT CTT CCT CTA TGT TG		
rpoC1 intron	5′*rpoC1* exon: GGT CTT CCT AGY TAY ATH GC	Downie, Katz-Downie, and Cho, 1996	AF031457–AF031462
	rpoC1 exon 2: ATT TCA TAT TCG AAY AAN CC		
AdhC	P1: CTG CKG TKG CAT GGG ARG CAG GGA AGC C	K. Schierenbeck, California State University, Fresno, personal communication (P1, P2); this paper (ADHX6-2, 4-1, and 8-2)	AF036567–AF036579
	P2: GCA CAG CCA CAC CCC AAC CCT G		
	ADHX6-2: TCA ATA CCA ATG ATC CTA GAA		
	ADHX4-1: TCA TGT TCT CCC TAT CTT CAC		
	ADHX8-2: GAA ACC ATG GCC TGG GTG		

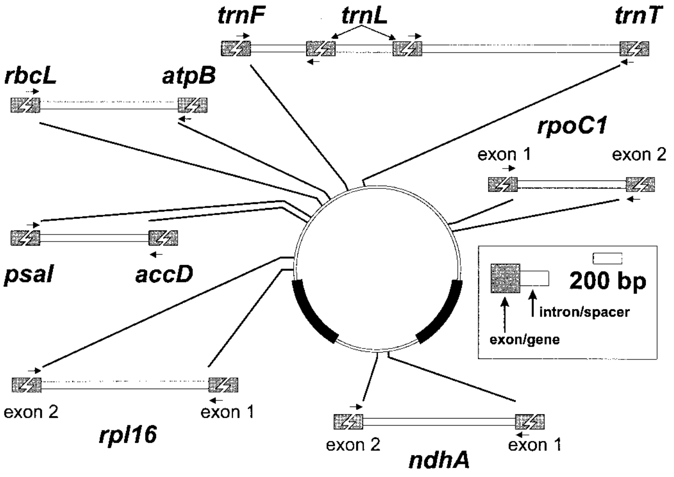

Fig. 2. Chloroplast DNA noncoding regions sampled. The circle represents the chloroplast genome, with shaded regions representing the inverted repeats. Sequenced regions are shown as mapped in the tobacco chloroplast genome (Shinozaki et al., 1986). For each region exons are represented by shaded boxes and are not drawn to scale; introns and spacers are represented by open boxes and are drawn approximately to scale.

Savolainen, Spichiger, and Manen, 1997). The *trnL-trnF* and *trnT-trnL* spacers were initially characterized by Taberlet et al. (1991). The *trnL-trnF* spacer has been widely exploited in molecular systematic investigations (Böhle et al., 1994; Gielly and Taberlet, 1994b; van Ham et al., 1994; Böhle, Hilger, and Martin, 1997; Sang, Crawford, and Stuessy, 1997). Curiously, the *trnT-trnL* spacer has rarely been used in systematic studies (Böhle et al., 1994; Böhle, Hilger, and Martin, 1997) despite the popularity of the other regions described in the same paper (Taberlet et al., 1991), the larger size of this region relative to the *trnL* intron and the *trnL-trnF* spacer, and the observation by Böhle et al. (1994) that this region is the most variable of the three. The *accD-psaI* spacer has been used only recently (Mendenhall, 1994; T. Barkman, University of Texas, Austin, personal communication). The PCR primers for the *accD-psaI* spacer region were originally designed by B. Milligan (New Mexico State University, Las Cruces) and were provided by T. Barkman and B. Simpson (University of Texas, Austin). The *ndhA* intron has been used in PCR-RFLP analysis (Wolf, Murray, and Sipes, 1997), and Downie, Katz-Downie, and Cho (1996) report 67.1% similarity in a comparison of the *ndhA* introns of tobacco and rice, but analyses of sequence variation among species have not previously been reported. PCR primers for the *ndhA* intron were designed based on maize, rice, and tobacco *ndhA* sequences from GenBank and were anchored in flanking exons. The *rpl16* intron has recently been used extensively for phylogenetic analyses in a variety of plant groups (Dickie, 1996; Jordan, Courtney, and Neigel, 1996; Kelchner and Wendel, 1996; Kelchner and Clark, 1997; Baum, Small, and Wendel, 1998; A. Schnabel and J. Wendel, unpublished data; S. Downie, University of Illinois, personal communication). Downie, Katz-Downie, and Cho (1996) report 64.5% similarity in a comparison of the *rpl16* introns of tobacco and rice; this is the lowest similarity reported in their comparison of cpDNA introns. The *rpoC1* intron was used by Downie, Katz-Downie, and Cho (1996) for assessing intrafamilial relationships within Apiaceae.

Nuclear-encoded alcohol dehydrogenase loci—Alcohol dehydrogenase (*Adh*, E.C. number 1.1.1.1) is a metabolic enzyme responsible for the interconversion of ethanol and acetaldehyde, primarily in response to hypoxic conditions (Freeling and Bennett, 1985). In cotton, as in most plants, *Adh* exists as a nuclear-encoded small gene family (Millar and Dennis, 1996; Small and Wendel, unpublished data). Gene structure

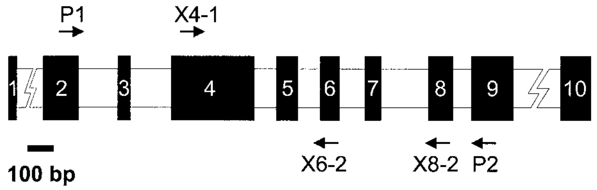

Fig. 3. Schematic representation of the *AdhC* genic region. Exons are represented by numbered and shaded boxes; introns are represented by open boxes. All regions are drawn to scale except introns 1 and 9 for which data are unavailable. Relative positions of the forward PCR primers are shown above the gene, and reverse primers are shown below.

of *Adh* in *Gossypium* is generally conserved relative to other plant species studied (Fig. 3; Millar and Dennis, 1996; Small and Wendel, unpublished data). Because the *Gossypium* species under consideration are allotetraploids (containing A and D subgenomes; see above) each nuclear-encoded locus present in diploid species is present in two copies (homoeologues) in the tetraploid species, one per subgenome. We have PCR-amplified, cloned, and sequenced the majority of a pair of homoeologous *Adh* genes from tetraploid *Gossypium* as well as the orthologues from diploid *Gossypium* representing the parents of the allopolyploid.

An underlying assumption of any phylogenetic analysis is that the sequences included are orthologous (related by speciation), rather than paralogous (related by gene duplication). The most reliable method of demonstrating orthology for nuclear genes is comparative genetic mapping. Mapping genes to positions on homologous/homoeologous linkage groups provides strong evidence for orthology. Therefore, we have genetically mapped the sequenced *Adh* loci in both the A- and D-diploid genomes and both subgenomes of the AD-allotetraploid. We found that these loci map to homologous/homoeologous linkage groups (data not shown) and so infer that they are orthologous. We term the *Gossypium* sequences reported here *AdhC* to differentiate them from the commonly used terminology *Adh1*, *Adh2*, etc., which imply homologies to *Adh* genes in other plants that are not in evidence. The *AdhC* sequences reported here are not orthologous to the *Gossypium Adh1* or *Adh2* sequences reported by Millar and Dennis (1996).

Adh sequences have been used previously in a number of phylogenetic and molecular evolutionary studies in plants (Gaut and Clegg, 1991, 1993; Goloubinoff, Pääbo, and Wilson, 1993; Hanfstingl et al., 1994; Gaut et al., 1996; Innan et al., 1996; Miyashita, Innan, and Terauchi, 1996; Morton, Gaut, and Clegg, 1996; Sang, Donoghue, and Zhang, 1997).

Amplification, cloning, and sequencing—*cpDNA regions*—PCR amplifications were performed in 50-μL reactions consisting of 1 unit *Taq* polymerase (Promega, Madison, Wisconsin), 1X buffer (Promega), 200 μmol/L each deoxy-nucleotide triphosphate, 1.5 mmol/L $MgCl_2$, 10–20 pmol of each primer, and 8–12 ng of template genomic DNA. Amplifications were carried out using the parameters described in Table 3 in an MJ Research PTC-100 thermal cycler (Watertown, Massachusetts). Amplifications were preceded by a "hotstart" consisting of 2 min at 94°C followed by 5 min at 80°C during which time the *Taq* polymerase was added to the reactions. A negative control reaction (no template DNA) was included for each set of amplifications to monitor for the possibility of contamination. All PCR primers were either obtained from other researchers or were synthesized by Integrated DNA Technologies (Coralville, Iowa). Amplification products were visualized by agarose gel electrophoresis, concentrated using Microcon-100 centrifugation separators (Amicon, Beverly, Massachusetts), and quantified fluorometrically. PCR products were either sequenced directly (*rpl16* intron, *trnL-trnF* spacer, *rpoC1* intron, *ndhA* intron) or cloned into pGEM-T (Promega) and sequenced (*atpB-rbcL* spacer, *trnT-trnL* spacer, *accD-psaI* spacer). For the cloning approach, purified PCR products were ligated into pGEM-T according to the manufacturer's instructions. Competent Top10 F′ (Invitrogen, San Diego, California) cells were transformed via electroporation and the resulting colonies were screened for plasmids with inserts by PCR using the original amplification primers. Plasmids were isolated from a single recombinant colony using an alkaline lysis/PEG precipitation protocol (Sambrook, Fritsch, and Maniatis, 1989). Cloning was performed only when PCR-amplification resulted in insufficient template for automated sequencing or when difficulties were encountered in using the amplification primers as sequencing primers. All sequencing was performed using amplification, internal, and/or vector specific primers (Table 2) at the Iowa State University DNA Sequencing and Synthesis Facility.

Adh—PCR-amplification and cloning of *Adh* homoeologues were performed as described for the cpDNA regions except that 2.0 mmol/L $MgCl_2$ was used in PCR reactions. The primers P1 and P2 (designed by K. Schierenbeck, California State University, Fresno; Table 2) are homologous to regions in exons 2 and 9, respectively, of *Gossypium Adh* (Fig. 3). Initial use of these primers resulted in amplification of multiple members of the *Adh* gene family. To isolate *AdhC* sequences, the entire heterogeneous PCR product pool was cleaned and concentrated using Geneclean II (Bio 101, La Jolla, California), ligated into pGEM-T, and transformed into Top10 F′ cells. The resulting colonies were screened by PCR using the amplification primers, and colonies that contained inserts of the size corresponding to the *AdhC* sequence were identified. Because tetraploid species of *Gossypium* contain two *AdhC* loci (homoeologues), it was necessary to further screen these plasmids to isolate A and D subgenome sequences. Multiple colonies containing plasmids with appropriately sized inserts were isolated from each taxon. Inserts from these plasmids were PCR amplified, ethanol precipitated, resuspended in a small volume of water, and then restriction digested with *Alu*I (American Allied Biochemical, Aurora, Colorado) according to the manufacturer's instructions. Visualization of the digestion products by agarose gel electrophoresis revealed subgenome-

Table 3. PCR amplification conditions.

Region	Denaturation temperature/Time	Annealing temperature/Time	Extension temperature/Time
atpB-rbcL spacer	94°C/1 min	55–50°C/1 min[a]	72°C/3 min
trnL-trnF spacer	94°C/1 min	48°C/1 min	72°C/3 min
trnT-trnL spacer	94°C/1 min	48°C/1 min	72°C/3 min
accD-psaI spacer	94°C/1 min	65°C/1 min 20 s	72°C/3 min
ndhA intron	94°C/1 min 30 s	42°C/1 min 30 s	72°C/2 min
rp116 intron	95°C/1 min	50°C/1 min[b]	65°C/4 min
rpoC1 intron	94°C/1 min	42°C/1 min	72°C/3 min
AdhC	94°C/1 min	50°C/1 min	72°C/2 min

[a] Touchdown PCR (Don et al., 1991); initial annealing temperature of 55°C, followed by a 0.5°C reduction in annealing temperature every cycle for ten cycles, followed by an additional 20 cycles with a 50°C annealing temperature.

[b] Following a 50°C annealing step for 1 min the temperature was ramped to 65°C by 1°/8 s.

Table 4. Characterization of cpDNA sequences (coding and noncoding).

Regions analyzed	Aligned length (bp)[a]	GC content	Divergence from A diploid outgroup[b]	Divergence within tetraploids[c]	Ts : Tv[d]	Substitutions[e]	Indels[f]
Intergenic spacers							
atpB-rbcL	976 (18)	28.3%	0.20%	0.20%	5:1	6(0)	5(0)
trnL-trnF	437 (42)	33.7%	0.12%	0.24%	1:2	3(0)	0
trnT-trnL	1394 (22)	22.9%	0.96%	0.49%	0.8:1	23(2)	6(2)
accD-psaI	1146 (390)	29.3%	0.40%	0.30%	2:1	12(0)	0
Introns							
ndhA	1140 (82)	31.9%	0.12%	0.04%	1:1	2(0)	0
rp116	1173 (24)	30.4%	0.34%	0.28%	0:6	6(2)	4(2)
rpoC1	1103 (353)	37.0%	0.00%	0.00%	—	0	0
Total	7369 (931)	30.0%	0.30%	0.20%	0.9:1	52(4)	15(4)

[a] Length of coding sequence in parentheses.

[b] Calculated as the mean nucleotide percentage difference between sequences from the outgroup (*G. arboreum*) and all ingroup species (gaps treated as missing data).

[c] Calculated as the mean nucleotide percent difference among all pairwise comparisons of sequences from tetraploid species (gaps treated as missing data).

[d] Ratio of transitions to transversions.

[e] Number of nucleotide substitutions; number of potentially phylogenetically informative substitutions in parentheses.

[f] Number of indels; number of potentially phylogenetically informative indels in parentheses.

specific digestion patterns that allowed discrimination of plasmids containing either A or D subgenome *AdhC* inserts. Using this PCR/cloning approach we isolated *AdhC* sequences from the diploids *G. robinsonii*, *G. raimondii* and from both the A and D subgenomes of all five tetraploid species (Table 1). These plasmids were then isolated and sequenced as described above. We were unable, however, to isolate the corresponding *AdhC* sequence from either of the two extant A-genome diploids (*G. arboreum* or *G. herbaceum*) using this approach. We therefore employed an internal, *AdhC*-specific primer (ADHX8–2, Table 2; Fig. 3) in conjunction with P1 and amplified a ~1.35 kb *AdhC* fragment from *G. arboreum*. Because the primer combination is locus specific we were able to directly sequence the *G. arboreum AdhC* PCR product using the Thermosequenase cycle-sequencing kit (Amersham, Arlington Heights, Illinois).

Analyses—Characterization of each region and sequence comparisons were facilitated by the software programs MacClade 3.05 (Sinauer, Sunderland, Massachusetts), PAUP 3.1.1 (Swofford, 1993) and MEGA 1.0 (Kumar, Tamura, and Nei, 1993). Analyses were conducted both on individual and combined data sets as follows. Individual cpDNA region data sets were analyzed separately (when warranted by the existence of sufficient variation) and then as a combined cpDNA data set. *Adh* sequences were analyzed in three separate ways: individual sequences as terminal "taxa," by subgenome, and by combining *Adh* homoeologue sequences for tetraploid taxa for an *Adh* "total evidence" analysis. For each data set a g_1 statistic (Hillis and Huelsenbeck, 1992; Hillis, Allard, and Miyamoto, 1993) was calculated using PAUP 3.1.1 to determine whether or not significant phylogenetic structure existed within the data set. For phylogenetic analyses, exhaustive searches for most-parsimonious trees were conducted with uninformative characters excluded. Due to the larger number of sequences included in the initial *Adh* analysis (each allotetraploid represented by two distinct sequences), the Branch and Bound algorithm was employed to search for maximally parsimonious trees. Relative levels of support for clades present in the most-parsimonious trees were assessed by calculating decay values, the number of extra steps required to collapse the clade (Bremer, 1988). For all phylogenetic analyses the tree lengths and consistency indices reported do not include autapomorphic characters. Rate variation among sequences was assessed using the 1D and 2D relative rate tests of Tajima (1993) as implemented in the program Tajima93 (T. Seelanan, unpublished software).

RESULTS

cpDNA sequences—Over 7.3 kilobase pairs (kb) of cpDNA sequence (6.4 kb of noncoding sequence) from seven different regions were determined for each of the five tetraploid species of *Gossypium* and the outgroup *G. arboreum*. These data collectively represent ~5.6% (7369/130 505 bp) of the unique sequence of the tobacco plastome (i.e., counting the inverted repeat only once) and ~10% (6438/64 437 bp) of the unique noncoding portion of the tobacco plastome (K. Wolfe, University of Dublin, Trinity College, Ireland, personal communication). Each of the sequenced regions is characterized in Table 4. Phylogenetically informative characters were observed only in the *trnT-trnL* spacer and the *rpl16* intron. The low observed GC content (~30%, see Table 4) of the sequenced regions is similar to that reported for plastomes in general (Palmer, 1991).

Averaged over all cpDNA sequences the mean divergence between *G. arboreum* and the ingroup species was 0.30% and mean divergence among tetraploid *Gossypium* was 0.20%. These values, however, were not equally distributed across all regions and, in fact, divergence from *G. arboreum* ranged from 0.00 to 0.96%, while divergence among tetraploids ranged from 0.00 to 0.49% (in both cases, *rpoC1* intron and *trnT-trnL* spacer, respectively). The mean transition:transversion ratio (Ts:Tv) across all cpDNA sequences was 0.9:1, while individual values ranged from 5:1 to 0:6 (Table 4). Substitution patterns taken across all regions appear to follow the observations of Morton (1995), in that positions flanked by A or T are more likely to undergo transversions. While this pattern is evident upon inspection, the data are too few to test statistically.

Overall, 7369 characters (nucleotides) were sampled, yielding 52 variable positions (0.71%) and four potentially phylogenetically informative nucleotide substitutions (0.05%). In addition to nucleotide substitutions, we observed 15 length mutations (indels), of which four were potentially phylogenetically informative.

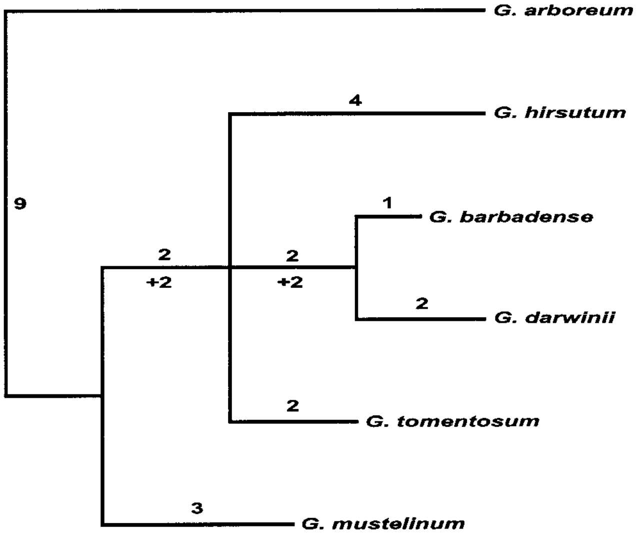

Fig. 4. Single most-parsimonious tree (length = 4, CI = 1.0, RI = 1.0) from analysis of the *trnT-trnL* spacer region. Branch lengths are shown above and decay values below each branch.

Phylogenetic analyses of cpDNA sequences—Potentially phylogenetically informative characters were found in only two of the seven regions: the *trnT-trnL* spacer (four characters) and the *rpl16* intron (four characters) (see Table 4). Exhaustive searches of all possible trees were performed for each of these data sets using PAUP v. 3.1.1 (Swofford, 1993). The g_1 statistics were -1.57 and -0.23 for the *trnT-trnL* and the *rpl16* intron, respectively. For the number of taxa and characters in these data sets, only the *trnT-trnL* spacer data set is significantly more structured than random ($P < 0.01$; Hillis and Huelsenbeck, 1992). The single most-parsimonious tree resulting from analysis of the *trnT-trnL* data set is shown in Fig. 4 (length = 4; consistency index [CI] = 1.0; retention index [RI] = 1.0). When all cpDNA data were combined into a single data set, a g_1 statistic of -1.08 was obtained which is significantly more structured than random ($P < 0.01$). Two equally most-parsimonious trees (length = 11; CI = 0.727; RI = 0.625) were found in an exhaustive search; the topology of the strict consensus tree was identical to Fig. 4. The two shortest trees differed only in the placement of *G. hirsutum,* which was resolved either as sister to a *G. barbadense* + *G. darwinii* clade or as part of an unresolved polytomy as in the strict consensus tree.

Nuclear Adh sequences—*Adh* exists as a small gene family in *Gossypium*. We chose to analyze the locus we refer to as *AdhC*. This locus maps to homologous/homoeologous regions of the A- and D-genome diploids and AD-genome tetraploid genetic maps (data not shown); thus we are confident that we are analyzing orthologous sequences. The PCR primers P1 and P2 amplify a ~1.65-kb region of *Adh* from exon 2 to the 5′ end of exon 9 (Fig. 3). We obtained sequences from *G. robinsonii* (C-genome diploid outgroup; see Fig. 1), *G. raimondii* (D-genome diploid), and from both the A- and D-subgenomes of all five AD-genome tetraploid species using these primers. A *G. arboreum* (A-genome diploid) sequence was obtained using the locus-specific primer pair P1/ADHX8-2, which amplifies a region from the middle of exon 2 to the 5′ end of exon 8 (Fig. 3); the resulting PCR product was 1352 bp in length.

All *AdhC* sequences maintain the expected 5′ GT... and ...AG 3′ intron boundary sequences with the exception of a G to A transition of the first nucleotide of intron 6 of the D-subgenomes of *G. hirsutum* and *G. tomentosum*, and an A to G transition at the 3′ end of intron 3. All sequences also maintain exon integrity (presence, length, reading frame) with the following exceptions. A 67-bp deletion in the A-subgenome sequences of *G. barbadense* and *G. darwinii* begins seven nucleotides from the 3′ end of exon 4 and ends in the middle of intron 4. A large (182 bp) deletion in the *G. arboreum* sequence results in partial loss of introns 5 and 6, and all of exon 6. Finally, a G to A transition in exon 2 of the *G. arboreum* sequence results in the conversion of a tryptophan-encoding codon (TGG) to a stop codon (TAG). The relevance of the foregoing observations to *AdhC* expression was not explored.

Sequence characteristics for *AdhC* are summarized in Table 5 and are discussed below. The total aligned length of the data matrix is 1667 bp; this includes 798 bp of exon sequence and 869 bp of intron sequence. With the exception of the sequence from *G. arboreum*, the absolute sequence lengths ranged from 1579 bp to 1655 bp. GC content varied little between the A- and D-(sub)genomes, but varied greatly between exons (45.4–46.2%) and introns (30.1–32.0%). Among sequences from tetraploid taxa, transition:transversion ratios (Ts:Tv) varied between genomes, and especially between introns and exons. In the A-(sub)genome the Ts:Tv was ~4.2:1, whereas in the D-(sub)genome the Ts:Tv was ~3.6:1 (Table 5). The differences between intron and exon Ts:Tv are more dramatic, ranging from 7–8:1 in exons to 1.6–3.3:1 in introns. Table 5 also reveals a marked disparity in the number of nucleotide substitutions in the two subgenomes; the number of nucleotide differences between all pairs of sequences are shown in Table 6. The D-subgenome sequences have experienced ~1.5 times as many nucleotide substitutions and yield almost three times as many potentially phylogenetically informative characters. This disparity is also reflected in the relative rate tests (Tajima, 1993), as summarized in Table 6. These tests indicate that, in all comparisons, *AdhC* genes from the D-(sub)genomes are accumulating substitutions at a rate that is significantly faster than are their orthologues/homoeologues in the A-(sub)genomes.

Phylogenetic analyses of Adh sequences—Three separate analyses were conducted with the *AdhC* sequences. First, an analysis was conducted using each sequence as a terminal; secondly, sequences of each (sub)genome were analyzed separately; and finally, the data from the subgenomes were combined for each taxon for a "total evidence" analysis.

For the data set in which each sequence was treated as a terminal the g_1 statistic estimated from 10 000 random trees was -0.49, which indicates that the data are significantly more structured than random ($P < 0.01$). Phylogenetic analysis of this data set resulted in a single most-parsimonious tree (length = 97, CI = 0.93, RI =

Table 5. Characterization of *Adh* sequences.

	Tetraploid taxa and diploid outgroups											Tetraploid taxa only		
Region analyzed	Aligned length (bp)	GC content (%)	Divergence from A/D diploid outgroup[a] (%)	Divergence from C diploid outgroup[b] (%)	Divergence within tetraploids[c] (%)	K_s[d]	K_a[d]	K[d]	Ts:Tv[e]	Substitutions[f]	Indels[g]	K_s[h]	K_a[h]	K[h]
A (sub)genome														
Exons	798	46.2	1.0	1.1	0.3	0.009	0.004	0.005	8:1	11(2)	2(1)	0.008	0.002	0.003
Introns	847	32.0	1.0	3.2	0.9	—	—	0.009	3.3:1	20(5)	2(0)	—	—	0.009
Total	1645	39.0	1.0	2.1	0.6	—	—	0.007	4.2:1	31(7)	4(1)	—	—	0.006
D (sub)genome														
Exons	798	45.4	1.9	2.3	1.4	0.028	0.013	0.016	7:1	33(12)	1(0)	0.019	0.013	0.014
Introns	865	30.1	2.6	5.5	0.7	—	—	0.014	1.6:1	30(6)	3(1)	—	—	0.008
Total	1663	37.5	2.3	3.9	1.1	—	—	0.015	3.6:1	63(18)	4(1)	—	—	0.011

[a] Calculated as the mean nucleotide percentage difference between the relevant subgenome outgroup (A—*G. arboreum* or D–*G. raimondii*) and the corresponding sequences from the tetraploid species (gaps treated as missing data).

[b] Calculated as the mean nucleotide percentage difference between the C-genome diploid outgroup (*G. robinsonii*) and sequences from the tetraploid species (gaps treated as missing data).

[c] Calculated as the mean nucleotide percentage difference among all pairwise comparisons of sequences from tetraploid species (gaps treated as missing data).

[d] Nucleotide substitutions among tetraploid taxa and their relevant diploid outgroup. Number of synonymous substitutions per synonymous site (K_s), nonsynonymous substitutions per nonsynonymous site (K_a), and substitutions per site (K) calculated with the Jukes and Cantor (1969) correction for multiple hits. K_s and K_a calculated according to the method of Nei and Gojobori (1986).

[e] Ratio of transitions to transversions among sequences from tetraploid taxa and the relevant subgenome outgroup.

[f] Number of nucleotide substitutions among sequences from tetraploid taxa and the relevant subgenome outgroup; number of potentially phylogenetically informative substitutions in parentheses.

[g] Number of indels among sequences from tetraploid taxa and the relevant subgenome outgroup; number of potentially phylogenetically informative indels in parentheses.

[h] Nucleotide substitutions among tetraploid taxa only. K_s, K_a, and K calculated with the Jukes and Cantor (1969) correction for multiple hits. K_s and K_a calculated according to the method of Nei and Gojobori (1986).

0.98), which is shown in Fig. 5. The tree is completely resolved and divided into two primary clades—one including the D-genome diploid and D-subgenome of the allotetraploids and the second including the A-genome diploid and the A-subgenomes of the allotetraploids. Within each (sub)genomic clade the resolution is complete and the topology is identical between clades.

Analyses of the subgenome sequences individually were also performed. The g_1 statistics calculated for the A- and D-subgenome data sets were -1.55 and -1.52, respectively; both values indicate data significantly more structured than random at the $P = 0.01$ level. In both cases, phylogenetic analysis found a single most-parsimonious tree. For the A-(sub)genome the tree had a length = 8, CI = 1.0, and RI = 1.0. The D-(sub)genome tree had a length = 20, CI = 0.95, and RI = 0.95. Again, each tree was fully resolved and the resulting topologies were identical to that shown in Fig. 5.

Finally, the data for both homoeologues were combined for each taxon for an *Adh* "total evidence" analysis. For outgroup comparison, the *G. raimondii* and *G. arboreum* sequences were combined to make a "diploid progenitor" sequence and the *G. robinsonii* sequence was duplicated. This data set had a g_1 statistic of -1.39, significantly more structured than random at the $P = 0.01$ level. An exhaustive search found a single most-parsi-

Table 6. Results of Tajima (1993) 2D relative rate tests for *Adh* sequences (below diagonal) and number of nucleotide differences between *Adh* sequences (above diagonal). Significantly different rates are denoted by asterisks as follows: * $0.05 > P > 0.01$; ** $0.01 > P > 0.005$; *** $P < 0.005$. A′ and D′ refer to the A and D subgenomes of the tetraploids. In all cases *G. robinsonii* was used as the reference taxon.

	Species											
Species	1	2	3	4	5	6	7	8	9	10	11	12
1 *G. raimondii* (D)	—	44	36	41	33	38	28	70	62	68	64	64
2 *G. hirsutum* D′		—	22	9	23	24	34	77	69	77	73	71
3 *G. barbadense* D′			—	19	17	4	28	71	63	71	67	65
4 *G. tomentosum* D′				—	20	21	31	76	68	76	72	70
5 *G. mustelinum* D′					—	19	29	68	60	68	64	62
6 *G. darwinii* D′						—	30	73	65	73	69	67
7 *G. arboreum* (A)	**	***	**	***	*	**	—	8	4	8	5	4
8 *G. hirsutum* A′	**	***	***	***	**	***		—	8	12	12	10
9 *G. barbadense* A′	**	***	***	***	***	***			—	13	9	2
10 *G. tomentosum* A′	**	***	***	***	*	***				—	14	15
11 *G. mustelinum* A′	**	***	***	***	**	***					—	11
12 *G. darwinii* A′	*	***	***	***	**	***						—

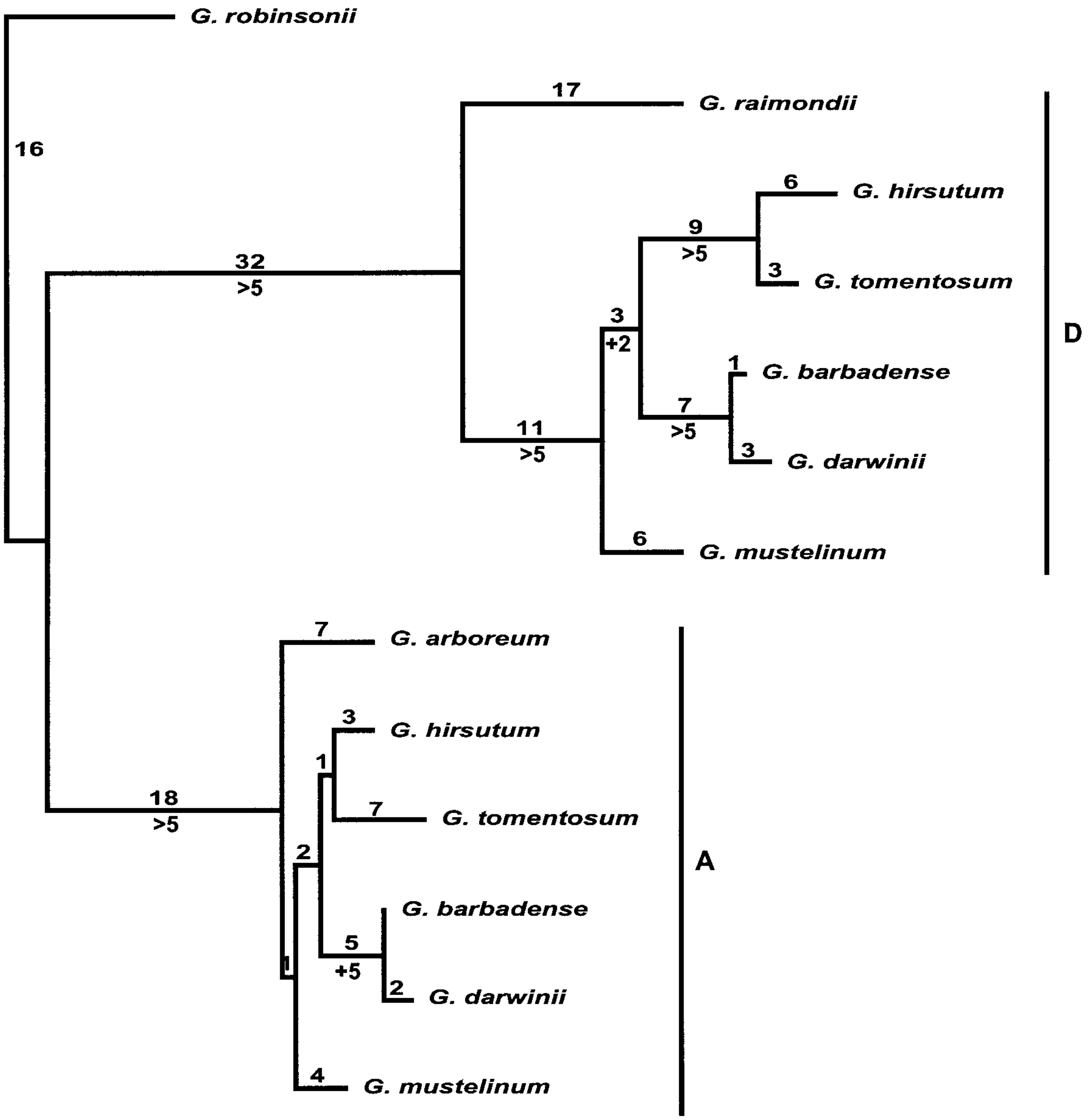

Fig. 5. Single most-parsimonious tree (length = 97, CI = 0.93, RI = 0.98) from analysis of individual *AdhC* sequences. Branch lengths are shown above and decay values below each branch. Nodes without decay values shown collapse in the strict consensus tree of trees one step longer than the most parsimonious.

monious tree (Fig. 6) with length = 43, CI = 0.91, and RI = 0.91. The tree is fully resolved and well supported, as indicated by high decay values and branch lengths.

DISCUSSION

Phylogeny of allotetraploid Gossypium—Despite intensive study of the tetraploid species of *Gossypium*, the phylogenetic relationships among these species have remained elusive. The data presented in this paper provide a completely resolved and robustly supported phylogenetic hypothesis for tetraploid *Gossypium* (Fig. 6). Within the tetraploid clade, the Brazilian endemic *G. mustelinum* represents the sole descendant of one branch of the initial divergence, as tentatively shown by DeJoode and Wendel (1992) and predicted by Wendel, Rowley, and Stewart (1994). The remaining four taxa form a clade sister to *G. mustelinum* and are divided into two species-pairs: *G. barbadense* + *G. darwinii* and *G. hirsutum* + *G. tomentosum*. The relationship between *G. barbadense* and *G. darwinii* has long been established, and in fact, the two taxa have been considered conspecific (see discussion in Percy and Wendel, 1990; Wendel and Percy, 1990). The affinities of *G. hirsutum* and *G. tomentosum*, however, were unclear until the study of DeJoode and Wendel (1992), which suggested that they are sister taxa; this relationship, however, was only weakly supported by a single rDNA restriction site mutation. Subsequent analysis of ITS sequences have confirmed this observation (Wendel, Schnabel, and Seelanan, 1995a, b; Seelanan,

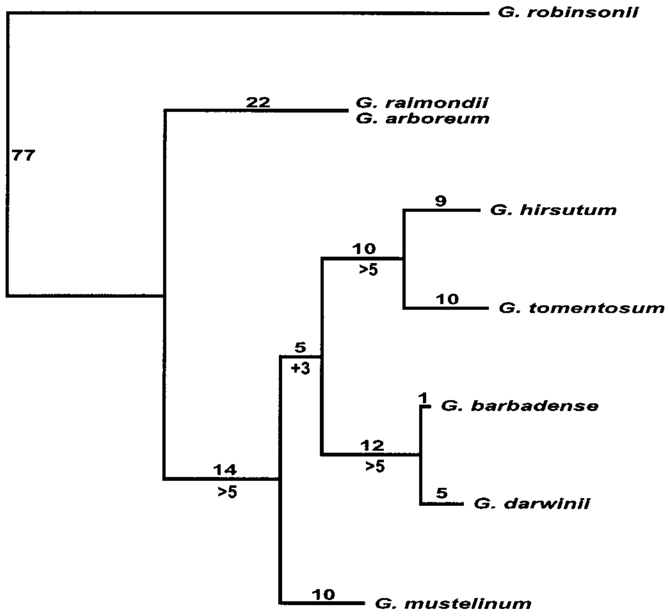

Fig. 6. Single most-parsimonious tree (length = 43, CI = 0.91, RI = 0.91) from analysis of combined *AdhC* data. Branch lengths are shown above and decay values below each branch.

Schnabel, and Wendel, 1997) and the *AdhC* data presented here corroborate this relationship and provide additional strong support.

Relationships hypothesized by these data additionally confirm predictions based on other sources of evidence. For example, the basal position of *G. mustelinum* predicts that it should be genetically equidistant from all other tetraploid species (Wendel, Rowley, and Stewart, 1994). This is borne out not only by the allozyme data presented by Wendel, Rowley, and Stewart (1994), but also by the *AdhC* sequence data reported in this paper; in the combined analysis (Fig. 6) there are 34, 35, 28, and 32 character-state changes between *G. mustelinum* and *G. hirsutum*, *G. tomentosum*, *G. barbadense* and *G. darwinii*, respectively (mean divergence from *G. mustelinum* = 1.0%). The *Adh* data also support the conclusion that *G. barbadense* and *G. darwinii* diverged more recently from each other than did *G. hirsutum* and *G. tomentosum*: while the branches leading to these two clades have similar lengths (10 vs. 12 steps), the number of autapomorphies each lineage has accumulated differ dramatically (9 and 10, respectively, in *G. hirsutum* and *G. tomentosum* vs. 1 and 5, respectively, in *G. barbadense* and *G. darwinii*).

Molecular evolution of noncoding cpDNA—The impetus for the experiments described here was to explore the phylogenetic utility of various sequences rather than to provide an in-depth analysis of patterns of molecular evolution. Nonetheless, some observations are prompted by our data. First, it has been recognized that cpDNA accumulates nucleotide substitutions more slowly than does plant nuclear DNA (Wolfe, Li, and Sharp, 1987; Wolfe, Sharp, and Li, 1989). As summarized in Tables 4 and 6, this rate difference is clearly evident in our data. In fact, the cpDNA data are astounding in their lack of informativeness, with a total of only eight phylogenetically informative characters observed among over seven thousand nucleotides surveyed. As a result of so little variation, the cpDNA provide only limited phylogenetic power.

In addition to the overall paucity of genetic variation, certain patterns observed previously are also noted here. First, the finding of Morton (1995) that transversions are more prevalent at positions flanked by A/T is supported by our data qualitatively, but sufficient data do not exist to statistically test this association. Also, previous observations that indels occur almost as frequently as nucleotide substitutions in noncoding cpDNA (Golenberg et al., 1993; Gielly and Taberlet, 1994b) are not supported by our data (Table 4). Rather, we detected over three times as many substitutions as indels in sequences from the allopolyploids (52 vs. 15, Table 4). Patterns of substitutions and indels vary between regions and in no case does the number of indels equal the number of substitutions. Of the indels that occur, two primary types are observed: insertion/deletion of a multinucleotide stretch of unique sequence or insertion/deletion of one or a few nucleotides within a polynucleotide tract (particularly polyA/T). The former type of indel is generally easily aligned and, if cladistically informative, is usually nonhomoplasious. In our cpDNA data there were 12 such indels, of which three were phylogenetically informative and none were homoplasious. The latter type of indel (three in our data), however, appears evolutionarily labile and probably originates via slipped-strand mispairing during replication (Levinson and Gutman, 1987). These types of indels often provide homoplasious characters. For example, the single homoplasious indel character in our cpDNA data set is a deletion of a single T in a string of ten in the *rpl16* intron, which is shared by *G. hirsutum* and *G. barbadense*.

Molecular evolution of Adh—Patterns of molecular evolution among the *AdhC* sequences will be discussed in the context of a full presentation of the evolution of the *Adh* gene family in *Gossypium*. Certain features of the data, however, are especially relevant here. In particular, the disparity of substitution rates between *AdhC* sequences of the A- and D-subgenomes is striking, consistent, and statistically significant (see Table 6). Relative rate differences may be attributed to a number of evolutionary or population genetic phenomena, including background mutational processes, generation time, lineage effects, selection, drift, and rates of recombination (Bosquet et al., 1992; Gaut et al., 1992; Gaut, Muse, and Clegg, 1993; Clegg et al., 1994; Eyre-Walker and Gaut, 1997). Because both of the two *AdhC* homoeologues exist within the same nuclear genome, however, background mutational and population genetic phenomena should affect them equally and can therefore be ruled out as having a significant effect. Selection is one (but not the only) process that can potentially differentially affect genes in the same nucleus. Either differing levels of purifying selection on the subgenome sequences or positive (diversifying or directional) selection on the D-subgenome sequences could account for the observed rate differences. There is an almost fivefold elevation of nucleotide substitution rates in exons of the D-subgenome relative to

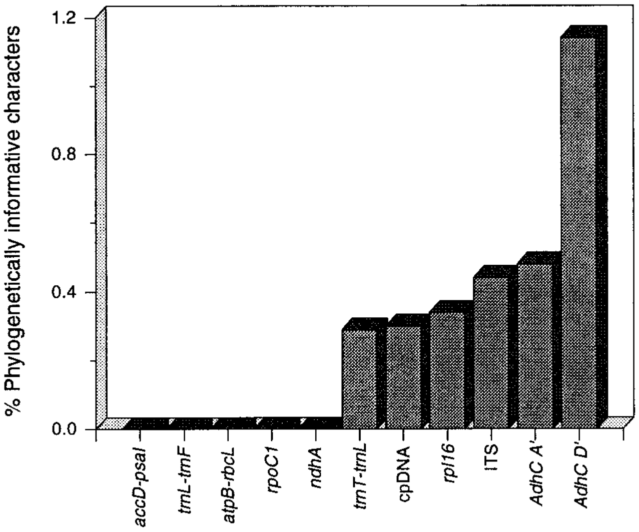

Fig. 7. Percentages of phylogenetically informative characters for several molecular data sets applied to tetraploid *Gossypium*. Number of informative ITS characters were partially extrapolated (see text).

the A-subgenome (K = 0.014 vs. 0.003, respectively; Table 5), despite the fact that intron nucleotide substitution rates are actually slightly higher in the A-subgenome sequences (K = 0.009 vs. 0.008; Table 5). Secondly, within exon sequences the synonymous nucleotide substitution rate (K_s) is over twice as high in the D-subgenome relative to the A-subgenome (K_s = 0.019 vs. 0.008; Table 5), but the nonsynonymous nucleotide substitution rate (K_a) is over six times higher (K_a = 0.013 vs. 0.002; Table 5). Finally, overall *AdhC* nucleotide substitution rates in the A-subgenome sequences are higher in the introns than in the exons (K = 0.009 vs. 0.003, respectively; Table 5) as predicted by neutral theory (Kimura, 1983); yet, in the D-subgenome sequences the nucleotide substitution rate is approximately twice as high in exons as in the introns (K = 0.014 vs. 0.008 respectively; Table 5). These data collectively suggest that selective forces may differ between homoeologues.

Relative phylogenetic utilities of molecular data—The phylogenetic conclusions described above are based almost exclusively on the wealth of data provided by the *AdhC* sequences, despite the volume of cpDNA data generated for identical taxa. In addition to the data presented in this paper, there exist for allotetraploid *Gossypium* comparable molecular data sets for cpDNA restriction sites (Wendel, 1989; DeJoode and Wendel, 1992; Wendel and Albert, 1992), and ITS sequences (Wendel, Schnabel, and Seelanan, 1995a, b; Seelanan, Schnabel, and Wendel, 1997). Figure 7 presents a comparison of the percentage of phylogenetically informative characters for these data sets. The cpDNA data consistently exhibit lower levels of informative characters than do the nuclear-encoded loci, as expected (Wolfe, Li, and Sharp, 1987; Wolfe, Sharp, and Li, 1989; Eyre-Walker and Gaut, 1997). The percentage of phylogenetically informative characters in the cpDNA data sets varied from 0 to 0.34%, and several of the cpDNA noncoding regions yielded no informative characters. The three cpDNA data sets that did contain informative characters (*rpl16* intron, *trnT-trnL* spacer, and cpDNA restriction sites) exhibited similar levels of informativeness both in terms of percentages (0.29–0.34%) and absolute numbers of informative characters (3–4).

Among the nuclear-encoded loci there is a large range of divergence values, as is expected given that each sequence type has its own unique biology. The value for ITS was partially extrapolated from the number of characters on internal branches. This was done because the ITS sequences in *G. mustelinum* have concerted to an A-genome like sequence, while the ITS sequences of the remaining tetraploids have concerted to a D-genome like sequence (Wendel, Schnabel, and Seelanan, 1995a). Despite these caveats, three results are clear from Fig. 7: (1) levels of phylogenetically informative characters are higher in nuclear-encoded sequences than in plastome data sets; (2) levels of informative characters vary among nuclear-encoded sequences; and (3) percentages of informative characters in the *AdhC* sequences are equivalent to or higher than ITS sequences. Current work is underway to examine levels of divergence among a large number of nuclear-encoded sequences in *Gossypium* (Cronn and Wendel, unpublished data).

Advantages and limitations of nuclear-encoded genes for phylogenetic analysis—*Relative rates*—It has long been recognized that nuclear-encoded sequences evolve at a faster rate than plastid-encoded sequences (e.g., Wolfe, Li, and Sharp, 1987; Wolfe, Sharp and Li, 1989; Eyre-Walker and Gaut, 1997). Despite this, in the search for the most phylogenetic information per unit of effort, nuclear-encoded sequences have been relatively ignored, with the exception of the widely used rDNA regions. The data presented here show clearly that cpDNA noncoding sequences may not be able to provide sufficient characters for robust resolution among closely related taxa, even if sampled ad infinitum. We sampled over 6 kb of cpDNA noncoding sequence (~10% of all unique cpDNA noncoding sequences) and yet obtained incomplete and poorly supported phylogenetic resolution. In addition, over 1000 cpDNA restriction sites were previously sampled (Wendel, 1989; DeJoode and Wendel, 1992), again with incomplete resolution. In contrast, sequences from a 1.6-kb nuclear-encoded *AdhC* gene provided complete and robust resolution among these closely related taxa. This difference in phylogenetic utility reflects simply the greatly accelerated rates of nucleotide substitution in the nuclear genome relative to the plastome, as illustrated in Fig. 7. The mean number of substitutions per site (K) in the combined cpDNA sequence data set was K = 0.002, while in the *AdhC* data sets K = 0.006 in the A-(sub)genome and K = 0.011 in the D-(sub)genome—a three to sixfold difference in nucleotide substitution rates. Extrapolation of these data allows the following observation. Given a total of four informative nucleotide substitutions out of a total of 6438 bp of noncoding cpDNA sequenced, and 25 informative nucleotide substitutions in the *AdhC* sequences, and assuming that levels of informative characters are constant across the chloroplast genome, over 40 kb of noncoding cpDNA would have to be sequenced to obtain an equivalent number of informative nucleotide substitutions as found in the *AdhC* sequences. This represents 62% (40 238 bp/64 437 bp) of the unique noncoding complement of the tobacco chloroplast genome (K. Wolfe, University of Dublin, Trinity College, Ireland, personal communication).

Patterns of mutation—In addition to levels of divergence, issues of alignability are important in selecting a genic or noncoding region for phylogenetic studies. While noncoding sequences generally accumulate nucleotide substitutions at a higher rate than coding sequences, they also appear to accumulate indels at a faster rate, occasionally equaling the rate of nucleotide substitutions (Golenberg et al., 1993; Gielly and Taberlet, 1994b). Because coding regions are constrained to maintain frame, indels occur less frequently, and when they do, they occur in multiples of three (i.e., a codon). Sequence alignment for genic regions, therefore, is usually straightforward, thereby making assessment of positional homology unambiguous. Noncoding regions, on the other hand, experience indel mutations of all lengths and at high frequency, making sequence alignment more problematic in many cases, particularly as more distantly related taxa are included (e.g., Golenberg et al., 1993; Downie, Katz-Downie, and Cho, 1996; Savolainen, Spichiger, and Manen, 1997). Additional confounding factors in assessing homology of mutations include the duplication/deletion of short repeats (or individual nucleotides in a run) via slipped-strand mispairing (Levinson and Gutman, 1987; Golenberg et al., 1993; Cummings, King, and Kellogg, 1994); the potential multiple origin of small inversions that occur in the loop of stem-loop secondary structures (Kelchner and Wendel, 1996); the higher potential for homoplasy due to a functionally reduced number of character states (due to the high AT content of noncoding cpDNA regions), and biased nucleotide substitutions in AT-rich regions (Morton, 1995). The use of coding regions can circumvent these difficulties, but at the cost of reduced levels of variation, at least in cpDNA genes. Nuclear-encoded genes, however, may offer the higher levels of variation desired, with the ease of alignment afforded by coding sequences.

Sequencing vs. restriction site data—Jansen, Wee, and Millie (1998) have analyzed both the relative utility (in terms of number of characters) and the relative reliability (in terms of CI and RI) of gene sequencing and restriction site studies of cpDNA. They suggest that, for intrageneric comparisons, cpDNA restriction site data are preferable, both because of the greater number of informative characters and because they report that restriction site data are, in general, less homoplasious than sequence data. Their analyses, however, did not address the lower end of the divergence spectrum (as in our study), where analysis of over 1000 cpDNA restriction sites still provided only limited resolution. cpDNA restriction site data are relatively free from problems associated with sequence data such as alignability. Comparison of mapped restriction sites is straightforward (assuming low levels of rearrangement), but becomes more difficult as taxonomic distance increases (Olmstead and Palmer, 1994; Jansen, Wee, and Millie, 1998). Restriction site studies, however, require large amounts of clean DNA and hence, are contraindicated in situations where availability of material is limiting.

Coalescence and intraspecific variation—Intraspecific genetic variation (i.e., allelic variation) is often observed when more than one accession of a species is sampled for molecular phylogenetic analysis. Two types of variation may be observed and their impacts on phylogenetic reconstruction are profoundly different. First, alleles within species may all be derived from a single ancestral allele present in the species, i.e., alleles coalesce within species. In this case, all intraspecific variation will be autapomorphic and therefore irrelevant for parsimony analysis. On the other hand, allelic variation may transcend species boundaries and therefore gene trees may not be equivalent to species trees simply because alleles may be older than species and multiple alleles can be maintained within a lineage (Pamilo and Nei, 1988; Hudson, 1990; Maddison, 1995; Clegg, 1997; Wendel and Doyle, 1998). The probability of concordance between a species tree and a gene tree is dependent on the time (in generations) between speciation events (the greater the number of generations, the higher the probability of recovering the species tree) and population genetic factors such as effective population size and selection. Although phylogenetic analyses of nuclear-encoded genes that have sampled multiple alleles are rare (see Huttley et al., 1997; Clegg, 1997, and references therein), incomplete coalescence has been observed (Buckler and Holtsford, 1996a, b; Gaut and Clegg, 1993; Goloubinoff, Pääbo, and Wilson, 1993; Hanson et al., 1996). Problems of noncoalescence are expected to be most prevalent in species where population genetic parameters promote the maintenance of multiple alleles, for example, large population size, high migration, and outbreeding (Pamilo and Nei, 1988; Hudson, 1990; Maddison, 1995). Populations of *Gossypium* species are primarily small, isolated, and inbred. These observations, in concert with the concordance of the phylogenies estimated from the separate homoeologues and the congruence with previous analyses, suggest to us that lack of coalescence is not an issue for this locus for these taxa. Current studies are underway to assess intraspecific polymorphism and to explicitly test whether or not *Adh* loci coalesce within closely related *Gossypium* species.

Concerted evolution—Multigene families are often subject to concerted evolution (Arnheim, 1983; Nagylaki, 1984; Walsh, 1987; Sanderson and Doyle, 1992; Elder and Turner, 1995). The ITS regions of nuclear rDNA became widely used as a source of sequence data after it became apparent that concerted evolution homogenizes sequences so that an entire array of tandemly repeated rDNA cistrons evolves as a single "locus" (Arnheim, 1983; Hillis and Dixon, 1991; Elder and Turner, 1995). Exceptions to the apparent rule of intraspecific and intraindividual sequence homogeneity are being discovered with increasing frequency, however, and the implications of these findings can be profound for phylogenetic reconstruction. Three observations that bear on the use of ITS are: (1) paralogous loci are not necessarily homogenized by concerted evolution (e.g., Suh et al., 1993); (2) in polyploids, interlocus concerted evolution may serve to homogenize homoeologous rDNA loci so that only a single parental type is retained, and this may occur differentially toward either parental type in different descendant lineages (Wendel, Schnabel, and Seelanan, 1995b; but see Waters and Schaal, 1996); and (3) rDNA pseudogenes may persist within the genome and may be preferentially sampled by PCR (Buckler and Holtsford, 1996a, b; Buckler, Ippolito, and Holtsford, 1997; Seelanan and Wendel, unpublished data). All three of the above phenomena may give rise to incongruence between the gene tree and the organismal tree, despite a well-resolved and robustly supported gene tree.

While interlocus gene conversion and recombination have been observed for low-copy nuclear-encoded gene families in plants (e.g., actins, Moniz de Sá and Drouin, 1996; heat-shock proteins, Waters, 1995; *rbcS*, Meagher, Berry-Lowe, and Rice, 1989; glutamine synthetase, Walker et al., 1995) the frequency of these events may depend on sequence conservation between paralogues (e.g., Walsh, 1987). Clearly, gene families that retain a large number of loci with strong sequence homologies are more likely to undergo interlocus concerted evolution and/or recombination than are smaller, more divergent gene families.

In our Southern hybridization experiments we used an *AdhC*-specific probe under high stringency conditions (65°C, 0.1 × SSC/0.5% SDS wash) and detected a single hybridizing band with multiple enzyme digestions for diploid taxa (data not shown) with the exception of *G. raimondii* (which showed a multibanded digestion pattern), and two hybridizing bands in the tetraploids. These Southern hybridization data, the recovery of two identical, paralogous gene trees, the genetic mapping data, and the high degree of sequence divergence between *Gossypium Adh* loci (16–25% in exons, introns are unalignable, Small and Wendel, unpublished data) provide strong evidence that homoeologues were sampled in the allotetraploids and that these sequences have been free from interlocus concerted evolution.

Conclusions—For phylogenetic analysis to accurately reconstruct organismal history (i.e., the species tree), orthologous sequences need to be compared (Wendel and Doyle, 1998). For this reason, among others, plant molecular systematics have relied primarily on cpDNA data because the chloroplast genome is nonrecombinant, generally uniparentally inherited, and “single copy.” Because nuclear-encoded genes usually exist in gene families, each member of which exists in a minimum of two copies (in diploids), and because these multiple copies may experience recombination and gene conversion, demonstration of orthology is more complex. Methods for establishing orthology (whether explicitly stated or implied) vary considerably and include criteria such as overall sequence similarity; monophyly and systematic content, i.e., reconstruction of the expected phylogeny (Gaut et al., 1996); tissue specificity (Doyle, 1991); Southern hybridization data (Matthews and Sharrock, 1996); and most convincingly, comparative genetic mapping data (Zhu et al., 1995; Cronn and Wendel, in press; this paper). These data are not always available or readily obtainable, but inferences of orthology may be facilitated with only a modest investment of effort by Southern hybridization experiments conducted using locus-specific probes and multiple enzyme digestions.

By isolating and analyzing orthologous nuclear genes and a number of different cpDNA regions, we have shown that mutation rates in noncoding cpDNA do not appear high enough to provide sufficient phylogenetic information to resolve relationships of this recently radiated group of tetraploid cottons, despite sequencing over 6 kb of noncoding cpDNA. Consequently, it is difficult to draw conclusions regarding the relative utility of the various cpDNA noncoding regions used. It is clear, however, that levels of divergence vary among noncoding cpDNA sequences (as pointed out for cpDNA introns by Downie, Katz-Downie, and Cho, 1996) and our analyses tentatively identify the *rpl16* intron and the *trnT-trnL* intergenic spacer as among the fastest evolving cpDNA regions (Table 4); this agrees with Downie, Katz-Downie, and Cho (1996), who suggested that *rpl16* should be the fastest evolving cpDNA intron.

As an alternative source of phylogenetic evidence, orthologous, low-copy, nuclear-encoded loci such as *AdhC* in *Gossypium*, may be isolated and may exhibit mutation rates up to six times higher than cpDNA noncoding sequences (Fig. 7). The use of nuclear-encoded genes for phylogeny reconstruction has both advantages and limitations. Primary among the advantages are the higher mutation rates and the ability to analyze large regions of sequence with interspersed coding and noncoding regions. The limitations, however, need to be considered. Demonstration of orthology among sequences is imperative and requires additional experimental effort. In addition, cognizance of issues such as coalescence and concerted evolution are required even when strict orthologues are recovered. Our study provides reason for both encouragement and caution in the continuing quest for additional and more informative tools for phylogenetic analysis in plants.

LITERATURE CITED

ARNHEIM, N. 1983. Concerted evolution of multigene families. *In* M. Nei and R. K. Koehn [eds.], Evolution of genes and proteins, 38–61. Sinauer, Sunderland, MA.

BAUM, D. A., R. SMALL, AND J. F. WENDEL. 1998. Biogeography and floral evolution of Baobabs (*Adansonia*, Bombacaceae) as inferred from multiple data sets. *Systematic Biology* 47: 181–207.

BAYER, R. J., L. HUFFORD, AND D. E. SOLTIS. 1996. Phylogenetic relationships in Sarraceniaceae based on rbcL and ITS sequences. *Systematic Botany* 21: 121–134.

BÖHLE, U.-R., H. HILGER, R. CERFF, AND W. F. MARTIN. 1994. Noncoding chloroplast DNA for plant molecular systematics at the infrageneric level. *In* B. Schierwater, B. Streit, G. P. Wagner, and R. DeSalle [eds.], Molecular ecology and evolution: approaches and applications, 391–403. Birkhäuser Verlag, Basel.

———, ———, AND W. F. MARTIN. 1997. Island colonization and evolution of the insular woody habit in *Echium* L. (Boraginaceae). *Proceedings of the National Academy of Sciences, USA* 93: 11740–11745.

BOSQUET, J., S. H. STRAUSS, A. H. DOERKSEN, AND R. A. PRICE. 1992. Extensive variation in evolutionary rate of *rbcL* gene sequences among seed plants. *Proceedings of the National Academy of Sciences, USA* 89: 7844–7848.

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

BUCKLER, E. S., AND T. P. HOLTSFORD. 1996a. *Zea* systematics: ribosomal ITS evidence. *Molecular Biology and Evolution* 13: 612–622.

———, AND ———. 1996b. *Zea* ribosomal repeat evolution and substitution patterns. *Molecular Biology and Evolution* 13: 623–632.

———, A. IPPOLITO, AND T. P. HOLTSFORD. 1997. The evolution of ribosomal DNA: divergent paralogs and phylogenetic implications. *Genetics* 145: 821–832.

CHASE, M. W., ET AL. 1993. Phylogenetics of seed plants: an analysis of nucleotide sequences from the plastid gene *rbc*L. *Annals of the Missouri Botanical Garden* 80: 528–580.

CLEGG, M. T. 1997. Plant genetic diversity and the struggle to measure selection. *Journal of Heredity* 88: 1–7.

———, B. S. GAUT, G. H. LEARN, JR., AND B. R. MORTON. 1994. Rates and patterns of chloroplast DNA evolution. *Proceedings of the National Academy of Sciences, USA* 91: 6795–6801.

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. *Journal of Molecular Evolution* 42: 685–705.

———, AND J. F. WENDEL. In press. Simple methods for isolating homoeologous loci from allopolyploid genomes. *Genome*.

CUMMINGS, M. P., L. M. KING, AND E. A. KELLOGG. 1994. Slipped-strand mispairing in a plastid gene: *rpo*C2 in grasses (Poaceae). *Molecular Biology and Evolution* 11: 1–8.

DEJOODE, D. R., AND J. F. WENDEL. 1992. Genetic diversity and origin of the Hawaiian islands cotton, Gossypium tomentosum. *American Journal of Botany* 79: 1311–1319.

DEMESURE, B., B. COMPS, AND R. J. PETIT. 1996. Chloroplast DNA phylogeography of the common beech (*Fagus sylvatica* L.) in Europe. *Evolution* 50: 2515–2520.

DICKIE, S. L. 1996. Phylogeny and evolution in the Subfamily Opuntioideae (Cactaceae): insights from *rpl*16 intron sequence variation. Master's thesis, Iowa State University, Ames, IA.

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 Research* 19: 4008.

DOWNIE, S. R., D. S. KATZ-DOWNIE, AND K.-J. CHO. 1996. Phylogenetic analysis of Apiaceae subfamily Apioideae using nucleotide sequences from the chloroplast *rpo*C1 intron. *Molecular Phylogenetics and Evolution* 6: 1–18.

DOYLE, J. J. 1991. Evolution of higher-plant glutamine synthetase genes: tissue specificity as a criterion for predicting orthology. *Molecular Biology and Evolution* 8: 366–377.

ELDER, J. F., JR., AND B. J. TURNER. 1995. Concerted evolution of repetitive DNA sequences in eukaryotes. *Quarterly Review of Biology* 70: 297–320.

ENDRIZZI, J. E., E. L. TURCOTTE, AND R. J. KOHEL. 1985. Genetics, cytology, and evolution of *Gossypium*. *Advances in Genetics* 23: 271–375.

EYRE-WALKER, A., AND B. S. GAUT. 1997. Correlated rates of synonymous site evolution across plant genomes. *Molecular Biology and Evolution* 14: 455–460.

FREELING, M., AND D. C. BENNETT. 1985. Maize *Adh1*. *Annual Review of Genetics* 19: 297–323.

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

GAUT, B. S., AND M. T. CLEGG. 1991. Molecular evolution of alcohol dehydrogenase 1 in members of the grass family. *Proceedings of the National Academy of Sciences, USA* 88: 2060–2064.

———, AND ———. 1993. Molecular evolution of the *Adh1* locus in the genus *Zea*. *Proceedings of the National Academy of Sciences, USA* 90: 5095–5099.

———, B. R. MORTON, B. C. MCCAIG, AND M. T. CLEGG. 1996. Substitution rate comparisons between grasses and palms: synonymous rate differences at the nuclear gene *Adh* parallel rate differences at the plastid gene *rbcL*. *Proceedings of the National Academy of Sciences, USA* 93: 10274–10279.

———, S. V. MUSE, W. D. CLARK, AND M. T. CLEGG. 1992. Relative rates of nucleotide substitution at the *rbc*L locus of monocotyledonous plants. *Journal of Molecular Evolution* 35: 292–303.

———, ———, AND M. T. CLEGG. 1993. Relative rates of nucleotide substitution in the chloroplast genome. *Molecular Phylogenetics and Evolution* 2: 89–96.

GIELLY, L., AND P. TABERLET. 1994a. Chloroplast DNA polymorphism at the intrageneric level and plant phylogenies. *Comptes Rendus des Seances, Academie des Sciences (Paris); Serie III Sciences de la vie/Life sciences* 317: 685–692.

———, AND ———. 1994b. The use of chloroplast DNA to resolve plant phylogenies: noncoding versus *rbc*L sequences. *Molecular Biology and Evolution* 11: 769–777.

———, AND ———. 1996. A phylogeny of the European gentians inferred from chloroplast *trn*L (UAA) intron sequences. *Botanical Journal of the Linnean Society* 120: 57–75.

———, Y.-M. YUAN, P. KÜPFER, AND P. TABERLET. 1996. Phylogenetic use of noncoding regions in the genus *Gentiana* L.: chloroplast *trn*L (UAA) intron versus nuclear ribosomal internal transcribed spacer sequences. *Molecular Phylogenetics and Evolution* 5: 460–466.

GOLENBERG, E. M., M. T. CLEGG, M. L. DURBIN, J. DOEBLEY, AND D. P. MA. 1993. Evolution of a noncoding region of the chloroplast genome. *Molecular Phylogenetics and Evolution* 2: 52–64.

GOLOUBINOFF, P., S. PÄÄBO, AND A. C. WILSON. 1993. Evolution of maize inferred from sequence diversity of an *Adh2* gene segment from archaeological specimens. *Proceedings of the National Academy of Sciences, USA* 90: 1997–2001.

HANFSTINGL, U., A. BERRY, E. A. KELLOGG, J. T. COSTA III, W. RÜDIGER, AND F. M. AUSUBEL. 1994. Haplotypic divergence coupled with lack of diversity at the *Arabidopsis thaliana* alcohol dehydrogenase locus: roles for both balancing and directional selection? *Genetics* 138: 811–828.

HANSON, M. A., B. S. GAUT, A. O. STEC, S. I. FUERSTENBERG, M. M. GOODMAN, E. H. COE, AND J. F. DOEBLEY. 1996. Evolution of anthocyanin biosynthesis in maize kernels: the role of regulatory and enzymatic loci. *Genetics* 143: 1395–1407.

HILLIS, D. M., M. W. ALLARD, AND M. M. MIYAMOTO. 1993. Analysis of DNA sequence data: phylogenetic inference. *Methods in Enzymology* 224: 456–487.

———, AND M. T. DIXON. 1991. Ribosomal DNA: molecular evolution and phylogenetic inference. *Quarterly Review of Biology* 66: 411–453.

———, AND J. P. HUELSENBECK. 1992. Signal, noise, and reliability in molecular phylogenetic analyses. *Journal of Heredity* 83: 189–195.

HODGES, S. A., AND M. L. ARNOLD. 1994. Columbines: a geographically widespread species flock. *Proceedings of the National Academy of Sciences, USA* 91: 5129–5132.

HUDSON, R. 1990. Gene genealogies and the coalescent process. *Oxford Surveys in Evolutionary Biology* 7: 1–44.

HUTTLEY, G. A., M. L. DURBIN, D. E. GLOVER, AND M. T. CLEGG. 1997. Nucleotide polymorphism in the chalcone synthase-A locus and evolution of the chalcone synthase multigene family of common morning glory *Ipomoea purpurea*. *Molecular Ecology* 6: 549–558.

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.

JANSEN, R. K., J. L. WEE, AND D. MILLIE. 1998. Comparative utility of chloroplast DNA restriction site and DNA sequence data for phylogenetic studies in plants. *In* D. Soltis, P. Soltis, and J. Doyle [eds.], Molecular systematics of plants, II: DNA sequencing, 87–100. Kluwer Academic Publishers, Boston, MA.

JOHNSON, D. A., AND J. HATTORI. 1996. Analysis of a hotspot for deletion formation within the intron of the chloroplast *trn*I gene. *Genome* 39: 999–1005.

JORDAN, W. C., M. W. COURTNEY, AND J. E. NEIGEL. 1996. Low levels of intraspecific genetic variation at a rapidly evolving chloroplast DNA locus in North American duckweeds (Lemnaceae). *American Journal of Botany* 83: 430–439.

JUKES, T. H., AND C. R. CANTOR. 1969. Evolution of protein molecules. *In* H. N. Munro [ed.], Mammalian protein metabolism, 21–132. Academic Press, New York, NY.

KELCHNER, S. A., AND L. G. CLARK. 1997. Molecular evolution and phylogenetic utility of the chloroplast *rpl16* intron in *Chusquea* and the Bambusoideae (Poaceae). *Molecular Phylogenetics and Evolution* 8: 385–397.

———, AND J. F. WENDEL. 1996. Hairpins create minute inversions in non-coding regions of chloroplast DNA. *Current Genetics* 30: 259–262.

KIMURA, M. 1983. The neutral theory of molecular evolution. Cambridge University Press, Cambridge.

Kita, Y., K. Ueda, and Y. Kadota. 1995. Molecular phylogeny and evolution of the Asian *Aconitum* subgenus *Aconitum* (Ranunculaceae). *Journal of Plant Research* 108: 429–442.

Kumar, S., K. Tamura, and M. Nei. 1993. MEGA: molecular evolutionary genetics analysis, version 1.0. The Pennsylvania State University, University Park, PA.

Levinson, G., and G. A. Gutman. 1987. Slipped-strand mispairing: a major mechanism for DNA sequence evolution. *Molecular Biology and Evolution* 4: 203–221.

Liston, A. 1992. Variation in the chloroplast genes *rpo*C1 and *rpo*C2 of the genus *Astragalus* (Fabaceae): evidence from restriction site mapping of a PCR-amplified fragment. *American Journal of Botany* 79: 953–961.

Maddison, W. 1995. Phylogenetic histories within and among species. *In* P. C. Hoch and A. G. Stephenson [eds.], Experimental and molecular approaches to plant biosystematics, *Monographs in Systematic Botany from the Missouri Botanical Garden*, vol. 53, 273–287.

Manen, J.-F., and A. Natali. 1995. Comparison of the evolution of ribulose-1, 5-bisphosphate carboxylase (*rbcL*) and *atpB-rbcL* noncoding spacer sequences in a recent plant group, the tribe Rubieae (Rubiaceae). *Journal of Molecular Evolution* 41: 920–927.

———, V. Savolainen, and P. Simon. 1994. The *atpB* and *rbcL* promoters in plastid DNAs of a wide dicot range. *Journal of Molecular Evolution* 38: 577–582.

Matthews, S., and R. A. Sharrock. 1996. The phytochrome gene family in grasses (Poaceae): a phylogeny and evidence that grasses have a subset of the loci found in dicot angiosperms. *Molecular Biology and Evolution* 13: 1141–1150.

Meagher, R. B., S. Berry-Lowe, and K. Rice. 1989. Molecular evolution of the small subunit of ribulose bisphosphate carboxylase: nucleotide substitution and gene conversion. *Genetics* 123: 845–863.

Mendenhall, M. 1994. Phylogeny of *Baptisia* and *Thermopsis* (Leguminosae) as inferred from chloroplast DNA and nuclear ribosomal DNA sequences, secondary chemistry, and morphology. Ph.D. dissertation, University of Texas, Austin, TX.

Millar, A. A., and E. S. Dennis. 1996. The alcohol dehydrogenase genes of cotton. *Plant Molecular Biology* 31: 897–904.

Miyashita, N. T., H. Innan, and R. Terauchi. 1996. Intra- and interspecific variation of the alcohol dehydrogenase locus region in wild plants *Arabis gemmifera* and *Arabidopsis thaliana*. *Molecular Biology and Evolution* 13: 433–436.

Moniz de Sá, M., and G. Drouin. 1996. Phylogeny and substitution rates of angiosperm actin genes. *Molecular Biology and Evolution* 13:1198–1212.

Morton, B. R. 1995. Neighboring base composition and transversion/transition bias in a comparison of rice and maize chloroplast noncoding regions. *Proceedings of the National Academy of Sciences, USA* 92: 9717–9721.

———, and M. T. Clegg. 1993. A chloroplast DNA mutational hotspot and gene conversion in a noncoding region near *rbcL* in the grass family (Poaceae). *Current Genetics* 24: 357–365.

———, B. S. Gaut, and M. T. Clegg. 1996. Evolution of alcohol dehydrogenase genes in the palm and grass families. *Proceedings of the National Academy of Sciences, USA* 93: 11735–11739.

Nagylaki, T. 1984. Evolution of multigene families under interchromosomal gene conversion. *Proceedings of the National Academy of Sciences, USA* 81: 3796–3800.

Natali, A., J.-F. Manen, and F. Ehrendorfer. 1995. Phylogeny of the Rubiaceae-Rubioideae, in particular the tribe Rubieae: evidence from a non-coding chloroplast DNA sequence. *Annals of the Missouri Botanical Garden* 82: 428–439.

Nei, M., and T. Gojobori. 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Molecular Biology and Evolution* 3: 418–426.

Olmstead, R. G., and J. D. Palmer. 1994. Chloroplast DNA systematics: a review of methods and data analysis. *American Journal of Botany* 81: 1205–1224.

———, and J. A. Sweere. 1994. Combining data in phylogenetic systematics: an empirical approach using three molecular data sets in the Solanaceae. *Systematic Biology* 43: 467–481.

Palmer, J. D. 1991. Plastid chromosomes: structure and evolution. *Cell Culture and Somatic Cell Genetics of Plants* 7A: 5–53.

Pamilo, P., and M. Nei. 1988. Relationships between gene trees and species trees. *Molecular Biology and Evolution* 5: 568–583.

Panero, J. L., and R. K. Jansen. 1997. Chloroplast DNA restriction site study of *Verbesina* (Asteraceae: Heliantheae). *American Journal of Botany* 84: 382–392.

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

Percy, R. G., and J. F. Wendel. 1990. Allozyme evidence for the origin and diversification of *Gossypium barbadense* L. *Theoretical and Applied Genetics* 79: 529–542.

Reinisch, A. J., J. Dong, C. L. Brubaker, D. M. Stelly, J. F. Wendel, and A. H. Paterson. 1994. A detailed RFLP map of cotton, *Gossypium hirsutum* × *Gossypium barbadense*: chromosome organization and evolution in a disomic polyploid genome. *Genetics* 138: 829–847.

Rieseberg, L. H., M. A. Hanson, and C. T. Philbrick. 1992. Androdioecy is derived from dioecy in Datiscaceae: evidence from restriction site mapping of PCR-amplified chloroplast DNA fragments. *Systematic Botany* 17: 324–336.

Sambrook, J., E. F. Fritsch, and T. Maniatis. 1989. Molecular cloning, a laboratory manual, 2d ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Sanderson, M. J., and J. J. Doyle. 1992. Reconstruction of organismal and gene phylogenies from data on multigene families: concerted evolution, homoplasy, and confidence. *Systematic Biology* 41: 4–17.

Sang, T., D. J. Crawford, and T. F. Stuessy. 1997. Chloroplast DNA phylogeny, reticulate evolution, and biogeography of *Paeonia* (Paeoniaceae). *American Journal of Botany* 84: 1120–1136.

———, M. J. Donoghue, and D. Zhang. 1997. Evolution of alcohol dehydrogenase genes in peonies (*Paeonia*): phylogenetic relationships of putative nonhybrid species. *Molecular Biology and Evolution* 14: 994–1007.

Savolainen, V., J. F. Manen, E. Douzery, and R. Spichiger. 1994. Molecular phylogeny of families related to Celastrales based on rbcL 5' flanking sequences. *Molecular Phylogenetics and Evolution* 3: 27–37.

———, R. Spichiger, and J.-F. Manen. 1997. Polyphyletism of Celastrales deduced from a chloroplast noncoding DNA region. *Molecular Phylogenetics and Evolution* 7: 145–157.

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

Shinozaki, K., et al. 1986. The complete nucleotide sequence of the tobacco chloroplast genome: its organization and expression. *EMBO Journal* 5: 2043–2049.

Soltis, D. E., P. S. Soltis, and J. J. Doyle. 1998. Molecular systematics of plants, II: DNA sequencing. Kluwer Academic Publishers, Boston, MA.

Soltis, P. S., D. E. Soltis, S. G. Weller, A. K. Sakai, and W. L. Wagner. 1996. Molecular phylogenetic analysis of the Hawaiian endemics *Schiedea* and *Alsinidendron* (Caryophyllaceae). *Systematic Botany* 21: 365–379.

Steele, K. P., and R. Vilgalys. 1994. Phylogenetic analyses of Polemoniaceae using nucleotide sequences of the plastid gene *mat*K. *Systematic Botany* 19: 126–142.

Suh, Y., L. B. Thien, H. E. Reeve, and E. A. Zimmer. 1993. Molecular evolution and phylogenetic implications of internal transcribed spacer sequences of ribosomal DNA in Winteraceae. *American Journal of Botany* 80: 1042–1055.

Swofford, D. L. 1993. PAUP: phylogenetic analysis using parsimony, version 3.1.1. Computer program distributed by the Illinois Natural History Survey, Champaign, IL.

Taberlet, P., L. Gielly, G. Pautou, and J. Bouvet. 1991. Universal primers for amplification of three non-coding regions of chloroplast DNA. *Plant Molecular Biology* 17: 1105–1109.

Tajima, F. 1993. Simple methods for testing the molecular evolutionary clock hypothesis. *Genetics* 135: 599–607.

van Ham, R. C. H. J., H. Hart, T. H. M. Mes, and J. M. Sandbrink. 1994. Molecular evolution of noncoding regions of the chloroplast genome in the Crassulaceae and related species. *Current Genetics* 25: 558–566.

WALKER, E. L., N. F. WEEDEN, C. B. TAYLOR, P. GREEN, AND G. M. CORUZZI. 1995. Molecular evolution of duplicate copies of genes encoding cytosolic glutamine synthetase in *Pisum sativum*. *Plant Molecular Biology* 29: 1111–1125.

WALSH, J. B. 1987. Sequence-dependent gene conversion: can duplicated genes diverge fast enough to escape conversion? *Genetics* 117: 544–557.

WATERS, E. R. 1995. The molecular evolution of the small heat-shock proteins in plants. *Genetics* 141: 785–795.

———, AND B. A. SCHAAL. 1996. Biased gene conversion is not occurring among rDNA repeats in the *Brassica* triangle. *Genome* 39: 150–154.

WENDEL, J. F. 1989. New World tetraploid cottons contain Old World cytoplasm. *Proceedings of the National Academy of Sciences, USA* 86: 4132–4136.

———. 1995. Cotton. *In* J. Smartt and N. W. Simmonds [eds.], Evolution of crop plants, 2d ed., 358–366. Longman Scientific & Technical, Essex.

———, AND V. A. ALBERT. 1992. Phylogenetics of the cotton genus (*Gossypium*): character-state weighted parsimony analysis of chloroplast-DNA restriction site data and its systematic and biogeographic implications. *Systematic Botany* 17: 115–143.

———, C. L. BRUBAKER, AND T. SEELANAN. In press. The origin and evolution of *Gossypium*. *In* J. Stewart, D. Oosterhuis, and J. Heitholt [eds.], Cotton physiology, Book II. Cotton Foundation, Memphis, TN.

———, AND J. J. DOYLE. 1998. Phylogenetic incongruence: window into genome history and molecular evolution. *In* D. Soltis, P. Soltis, and J. Doyle [eds.], Molecular systematics of plants, II: DNA sequencing, 265–296. Kluwer Academic Publishers, Boston, MA.

———, AND R. G. PERCY. 1990. Allozyme diversity and introgression in the Galapagos Islands endemic *Gossypium darwinii* and its relationship to continental *G. barbadense*. *Biochemical Systematics and Ecology* 18: 517–528.

———, R. ROWLEY, AND J. MCD. STEWART. 1994. Genetic diversity in and phylogenetic relationships of the Brazilian endemic cotton, *Gossypium mustelinum* (*Malvaceae*). *Plant Systematics and Evolution* 192: 49–59.

———, A. SCHNABEL, AND T. SEELANAN. 1995a. An unusual ribosomal DNA sequence from *Gossypium gossypioides* reveals ancient, cryptic, intergenomic introgression. *Molecular Phylogenetics and Evolution* 4: 298–313.

———, ———, AND ———. 1995b. Bidirectional interlocus concerted evolution following allopolyploid speciation in cotton (*Gossypium*). *Proceedings of the National Academy of Sciences, USA* 92: 280–284.

WOLF, P. G., R. A. MURRAY, AND S. D. SIPES. 1997. Species-independent, geographical structuring of chloroplast DNA haplotypes in a montane herb *Ipomopsis* (Polemoniaceae). *Molecular Ecology* 6: 283–291.

WOLFE, A. D., W. J. ELISENS, L. E. WATSON, AND C. W. DEPAMPHILIS. 1997. Using restriction-site variation of PCR-amplified cpDNA genes for phylogenetic analysis of Tribe Cheloneae (Scrophulariaceae). *American Journal of Botany* 84: 555–564.

WOLFE, K. H., W.-H. LI, AND P. M. SHARP. 1987. Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. *Proceedings of the National Academy of Sciences, USA* 84: 9054–9058.

———, P. M. SHARP, AND W.-H. LI. 1989. Rates of synonymous substitution in plant nuclear genes. *Journal of Molecular Evolution* 29: 208–211.

ZHU, T., L. SHI, J. J. DOYLE, AND P. KEIM. 1995. A single nuclear locus phylogeny of soybean based on DNA sequence. *Theoretical and Applied Genetics* 90: 991–999.