JOURNAL OF MOLECULAR EVOLUTION

Polymorphism and Concerted Evolution in a Tandemly Repeated Gene Family: 5S Ribosomal DNA in Diploid and Allopolyploid Cottons

Richard C. Cronn,[1] Xinping Zhao,[2] Andrew H. Paterson,[3] Jonathan F. Wendel[1]

[1]Department of Botany, Iowa State University, Ames, IA 50011, USA
[2]University of Michigan Medical Center, MSRB II, C568, Ann Arbor, MI 48109-0672, USA
[3]Department of Plant and Soil Sciences, Texas A & M University, College Station, TX 77843-2474, USA

Received: 27 June 1995 / Accepted: 6 November 1995

Abstract. 5S RNA genes and their nontranscribed spacers are tandemly repeated in plant genomes at one or more chromosomal loci. To facilitate an understanding of the forces that govern 5S rDNA evolution, copy-number estimation and DNA sequencing were conducted for a phylogenetically well-characterized set of 16 diploid species of cotton (*Gossypium*) and 4 species representing allopolyploid derivatives of the diploids. Copy number varies over twentyfold in the genus, from approximately 1,000 to 20,000 copies/2C genome. When superimposed on the organismal phylogeny, these data reveal examples of both array expansion and contraction. Across species, a mean of 12% of nucleotide positions are polymorphic *within* individual arrays, for both gene and spacer sequences. This shows, in conjunction with phylogenetic evidence for ancestral polymorphisms that survive speciation events, that intralocus concerted evolutionary forces are relatively weak and that the rate of interrepeat homogenization is approximately equal to the rate of speciation. Evidence presented also shows that duplicated 5S rDNA arrays in allopolyploids have retained their subgenomic identity since polyploid formation, thereby indicating that interlocus concerted evolution has not been an important factor in the evolution of these arrays. A descriptive model, one which incorporates the opposing forces of mutation and homogenization within a selective framework, is outlined to account for the empirical data presented. Weak homogenizing forces allow equivalent levels of sequence polymorphism to accumulate in the 5S gene and spacer sequences, but fixation of mutations is nearly prohibited in the 5S gene. As a consequence, fixed interspecific differences are statistically underrepresented for 5S genes. This result explains the apparent paradox that despite similar levels of gene and spacer diversity, phylogenetic analysis of spacer sequences yields highly resolved trees, whereas analyses based on 5S gene sequences do not.

Key words: 5S rDNA — Concerted evolution — *Gossypium* — Polyploidy — Molecular evolution — Genome evolution — Repetitive DNA — Molecular phylogeny

Introduction

A universal feature of ribosomes is the presence of small 5S RNA molecules that are associated with the large subunit. These RNAs usually are encoded by genes organized into tandemly repeated arrays that occur at one or more chromosomal loci (Long and Dawid 1980; Gerbi 1985; Appels and Honeycutt 1986; Sastri et al. 1992). At each locus, the 5S genes, which are approximately 120 bp in length, are separated from one another by intergenic, nontranscribed spacers, which in plants vary in length from approximately 100 to 700 bp. In plants, the total number of repeats (= gene + spacer) per genome varies by over two orders of magnitude, from less than 1,000 to over 100,000 (Schneeberger et al. 1989; Sastri et al. 1992).

Correspondence to: J.F. Wendel

Because of this ubiquity and prominence, 5S rDNA diversity and evolution have been studied in a broad range of plants with a particular focus on patterns of nucleotide conservation and divergence (Wolters and Erdmann 1988; Campell et al. 1992; Baum and Johnson 1994; Kellogg and Appels 1995) and on the structural organization of 5S arrays (Scoles et al. 1988; Dvorák et al. 1989; Schneeberger et al. 1989; Gottlob-McHugh 1990; Appels et al. 1992; Röder et al. 1992; Kanazin et al. 1993). This accumulating data base demonstrates that 5S RNA genes are highly conserved in the plant kingdom, both with respect to length and nucleotide sequence, whereas the intergenic spacers evolve more rapidly. An additional generalization is that individual 5S rDNA repeats do not evolve independently. As with 18S–26S rDNA and other tandemly repeated multigene families, the hundreds to thousands of repeats within 5S arrays retain a high degree of identity due to homogenizing forces collectively referred to as ''concerted evolution'' (Zimmer et al. 1980).

Most models of concerted evolution invoke one or two molecular processes, unequal crossing-over and gene conversion (Hood et al. 1975; Smith 1976; Dover 1982; Arnheim 1983; Ohta 1983, 1984, 1990; Ohta and Dover 1983; Nagylaki 1984a,b, 1990; Basten and Ohta 1992; Schlötterer and Tautz 1994). Regardless of mechanism, the rate at which variant repeat types become homogenized depends upon several factors, including the number of repeats in an array, the frequency of homogenization events relative to the formation of new variants via mutation, the intensity of natural selection, and effective population size (Smith 1976; Ohta 1983, 1984, 1990; Nagylaki 1984a,b, 1990; Li et al. 1985; Basten and Ohta 1992; Linares et al. 1994). Interplay among these and other variables leads to a continuum in the degree of heterogeneity exhibited by repeats within arrays: When concerted evolution is ''strong,'' repeats are identical or nearly so; when weaker, sequence heterogeneity is observed. With respect to 5S rDNA arrays, sequence heterogeneity among repeats within individual arrays has been reported from several plant groups (Rafalski et al. 1982; Gottlob-McHugh et al. 1990; Cox et al. 1992; Kellogg and Appels 1995). In each case, a moderate level of sequence variation has been detected, demonstrating that concerted evolutionary forces have not been strong enough to overcome those that generate variation.

An additional complexity arises when 5S arrays occur at more than one chromosomal location. This is the situation in the majority of plants, where polyploidy is prevalent (Masterson 1994). In these cases, the outcome of concerted evolutionary processes depends not only on the previously listed factors but also on the frequency of genetic exchanges between homologous sequences on homeologous chromosomes (Ohta and Dover 1983; Nagylaki 1984b 1990; Schlötterer and Tautz 1994). When barriers prevent such exchange, chromosome-specific arrays may evolve independently despite concerted evolution of repeats within arrays, as appears to be the case for 5S arrays in the grass tribe Triticeae (Scoles et al. 1988; Kellogg and Appels 1995). The alternative scenario, homogenization of repeats from different arrays (interlocus concerted evolution; e.g., Wendel et al. 1995a), has yet to be demonstrated for 5S rDNA in plants (Sastri et al. 1992).

As part of an ongoing effort to elucidate the evolutionary process in a model system involving diploid and allopolyploid members of *Gossypium* (VanderWiel et al. 1993; Reinisch et al. 1994; Wendel et al. 1995a,b), we here describe patterns of polymorphism in and concerted evolution of 5S rDNA sequences. Fluorescent in situ hybridization analysis has revealed that *Gossypium* 5S arrays occupy only a single centromeric chromosomal location in A-genome and D-genome diploid species (R. Hanson and D. Stelly, pers. comm) and two corresponding loci in the AD-genome allopolyploids (Crane et al. 1993). This organization contrasts with that of the major 18S–26S rDNA arrays, which occupy two relatively distal chromosomal loci in each diploid and in each subgenome of the allopolyploids. Repeats of *Gossypium* 18S–26S rDNA arrays evolve under strong inter- and intralocus concerted evolution (Wendel et al. 1995a), leading us to ask whether other tandemly repeated sequences evolve in a similar manner. To address this issue, we cloned and sequenced multiple 5S rDNA repeats from allopolyploid species of cotton and all lineages representing their diploid progenitor genomes. We were particularly interested in whether 5S sequences from individual arrays are homogeneous or polymorphic in diploid and allopolyploid species and in whether 5S sequences from the two arrays of the allopolyploid species evolve independently or in a concerted fashion.

Materials and Methods

Organismal Context. Gossypium is a genus of approximately 50 species that range from perennial herbs to small trees with centers of diversity in Australia, Africa-Arabia, and Mexico (Fryxell 1979, 1992). Extensive chromosomal diversification accompanied radiation of the genus, leading to the evolution of diploid ''genome groups,'' designated A–G based on chromosome size differences and meiotic pairing behavior in interspecific hybrids (reviewed in Endrizzi et al. 1985). Phylogenetic analyses based on chloroplast DNA restriction-site variation (DeJoode 1992; Wendel and Albert 1992) reveal diploid clades that are congruent with taxonomic alignments, geographic distributions, and genome designations. In addition to the diploid species, there are five species of tetraploid ($2n = 52$) *Gossypium* (Brubaker and Wendel 1993; Fryxell 1992), all endemic to the New World. A wealth of data establish that these tetraploid species are derived from allopolyploidization between A-genome and D-genome progenitors (Endrizzi et al. 1985; Wendel 1989; Reinisch et al. 1994). Chloroplast DNA sequence divergence data suggest that the two parental genomes diverged from a common ancestor 6–11 million years ago (MYA) and that they became reunited in a common nucleus in the mid-Pleistocene (1–2 MYA; Wendel 1989; Wendel and Albert 1992). Although the actual progenitor diploid taxa are most likely extinct, data suggest that the best models of the ances-

Table 1. Description of *Gossypium* accessions used[a]

Taxon	Genome designation	Accession	Geographic origin	Clones isolated	GenBank numbers
Subgenus *Sturtia* (R. Brown) Todaro					
Section *Sturtia*					
G. robinsonii F. von Mueller	C_2	AZ-50	Western Australia	3	U31852–U31854
Subgenus *Gossypium*					
Section *Gossypium*					
Subsection *Gossypium*					
G. herbaceum L.	A_1	A1-73	Botswana	5	U32006–U32010
G. arboreum L.	A_2	A2-74	China	5	U31855–U31856, U31999–U32001
Subgenus *Houzengenia* Fryxell					
Section *Houzingenia*					
Subsection *Houzingenia*					
G. trilobum (Mocino & Sesse ex DC) Skovsted	D_8	D8-1	western Mexico	3	U32056–U32058
G. thurberi Todaro	D_1	T17	Arizona, USA	3	U32059–U32061
Subsection *Integrifolia* (Todaro) Todaro					
G. davidsonii Kellogg	$D_{3\text{-d}}$	D3d-32a	Baja California, Mexico	2	U32054–D32055
G. klotzschianum Andersson	$D_{3\text{-k}}$	D3k-3	Galapagos Islands	4	U32050–U32053
Subsection *Caducibracteolata* Mauer					
G. armourianum Kearney	$D_{2\text{-}1}$	D2-1	Baja California, Mexico	2	U32072–U32073
G. harknessii Brandegee	$D_{2\text{-}2}$	D2-2	Baja California, Mexico	4	U32068–U32071
G. turneri Fryxell	D_{10}	D10-3	Sonora, Mexico	3	U32066–U32067, U39496
Section *Erioxylum* (Rose & Standley) Prokhanov					
Subsection *Erioxylum* (Rose & Standley) Fryxell					
G. aridum (Rose & Standley ex Rose) Skovsted	D_4	D4-12	Colima, Mexico	2	U32040–U32041
G. lobatum H. Gentry	D_7	D7	Michoacan, Mexico	5	U32042–U32046
G. laxum Phillips	D_9	LP	Guerrero, Mexico	3	U32047–U32049
G. schwendimanii Fryxell & Koch	—	JMS	Michoacan, Mexico	4	U32036–U32039
Subsection *Selera* (Ulbrich) Standley					
G. gossypioides (Ulbrich) Standley	D_6	D6-5	Oaxaca, Mexico	4	U32032–U32035
Subsection *Austroamericana* Fryxell					
G. raimondii Ulbrich	D_5	D5-37	Peru	6	U32074–U32077, U39497–U39498
Subgenus *Karpas* Rafinesque					
G. hirsutum L.	$(AD)_1$	Tx2094	Yucatán, Mexico	9	U32027–U32031, U32083–U32085, U39499
G. barbadense L.	$(AD)_2$	K101	Bolivia	7	U32011–U32017
G. tomentosum Nuttall ex Seemann	$(AD)_3$	WT936	Hawaii, USA	6	U32018–U32021, U39494–U39495
G. mustelinum Miers ex Watt	$(AD)_4$	JL	Ceará, Brazil	10	U32022–U32026, U32078–U32082
Doubled (*G. arboreum* × *G. thurberi*)	$2(A_2D_1)$	Beasley	Synthetic allopolyploid	9	U32002–U32005, U32062–U32065, U39491

[a] Taxa are arranged according to the classification of Fryxell (1992). Cytogenetic ("genome") designations follow the conventions of Endrizzi et al. (1985). Accession names are those used in the National Collection of *Gossypium* Germplasm (Percival 1987) or by our own laboratory. Geographic origin refers to site of accession collection rather than the aggregate range of the species. Between two and nine clones were isolated per species

tral D-genome parent are *G. raimondii* (Endrizzi et al. 1985) and *G. gossypioides* (Wendel and Albert 1992; Wendel et al. 1995b) and that the A-genome donor was most similar to present-day *G. herbaceum* (Endrizzi et al. 1985). Following polyploidization, allopolyploids diverged into five modern species (DeJoode and Wendel 1992; Brubaker and Wendel 1993; Wendel et al. 1994), including the commercially important *G. hirsutum* ("upland cotton") and *G. barbadense* ("Pima" and "Egyptian" cotton).

Plant Materials. We isolated total DNA, using methods detailed in Paterson et al. (1993), from individual plants representing 20 diploid and allopolyploid *Gossypium* species (Table 1). Included were both Old World A-genome diploids, all 13 New World D-genome diploids, and four of five New World AD-genome allopolyploids. Because Australian C-genome cottons are basal within the genus (Wendel and Albert 1992), we selected one Australian species (*G. robinsonii*) for inclusion as an outgroup. For comparative purposes, we also isolated DNAs from a synthetic allopolyploid, $2(A_2D_1)$, derived via colchicine-doubling of the sterile intergenomic F_1 hybrid *G. arboreum* × *G. thurberi* (synthesized by J. O. Beasley).

5S rDNA Amplification and Cloning. We amplified 5S repeats from each genomic DNA by the polymerase chain reaction (PCR), using the cycling parameters specified in Cox et al. (1992). Reactions were conducted in 50-μl volumes containing 10–500 ng of genomic DNA and 10 pmol of each primer. Primers 5SF (5′-GAG-AGT-AGT-AC[A/T]-[A/T][C/G]G-ATG-GG) and 5SR (5′-GGA-GTT-CTG-A[C/T]G-GGA-TCC-GG) were designed to anneal to the 5S rRNA gene at nucleotides 69–88 and 28–49, respectively. Under the PCR conditions of Cox et al. (1992) and a variety of other amplification protocols, products consisted of a ladder of 5S repeats that ranged from single repeats (~300 bp) to multimers greater than 10 kb in length. These ladders were digested with *Bam*HI to yield 5S pools that were mostly monomeric in length, and were readily cloned into M13mp18 (Gibco-BRL). Individual clones were isolated using routine procedures (Sambrook et al. 1989) and were sequenced using an ABI automated sequencer. A total of 99 5S rDNA sequences were generated for the 20 species and the synthetic allopolyploid $2(A_2D_1)$, with an average of four clones per diploid and eight clones per polyploid species. To evaluate sequencing error and to verify the intra-individual polymorphisms observed, residual DNAs from 40 clones were sequenced manually using standard methods of dideoxy sequencing.

5S rDNA Copy-Number Estimation. To quantify the number of 5S repeats in each species examined, DNAs were digested to completion with RNase A (to remove potentially contaminating RNAs) and subjected to slot-blot analysis. Previous studies (Edwards et al. 1974; Kadir 1976; Michaelson et al. 1991) demonstrate that 2C DNA content varies little among *Gossypium* species within genome groups, but more between species from different genome groups. For estimating slot-blot loadings, we used published 2C estimates (Edwards et al. 1974; Kadir 1976; Michaelson et al. 1991) that are in close agreement with newer data (J. S. Johnston and H. J. Price, pers. comm.): A genome = 3.8 pg; C genome = 5.0 pg; D genome = 2.0 pg; AD genome ≈ 5.8 pg.

Samples of genomic DNA, quantified by microfluorometry, were transferred onto MSI MagnaGraph nylon membranes using a 72-position slot-blot apparatus (Schleicher and Schuell). For each sample, 10^4 2C genomic equivalents of DNA were denatured in 0.4 M NaOH at 65C for 30 min, neutralized with an equal volume of 2 M NH_4OAc (pH 7), and transferred to the slot wells. Copy-number standards for A- and D-genome 5S rDNA were generated using PCR-amplified inserts from the M13 clones *G. herbaceum* pGh328-3 (A-genome) and *G. raimondii* pGr330-1 (D-genome). Quantified amounts (ranging from 10^6 to 10^9 copies) were applied in duplicate to each blot. DNAs were UV-crosslinked to membranes that were dried at 65C for 2 h. Hybridization probes were generated from both clones using random-primer labeling and the hybridization and wash conditions detailed in Wendel et al. (1995a). After washing, signal intensity was quantified by phosphorimaging on a Molecular Dynamics 400 phosphorimager. Absolute 5S rDNA copy numbers for each slot were estimated by linear regression using a standard curve of "volume above background" values (obtained from ImageQuant software). Copy numbers per 2C genome were calculated by dividing absolute number of copies per slot by the number of genomic equivalents loaded in that slot. To estimate standard errors for 5S copy number, four to eight replicate experiments were conducted.

Sequence Alignment and Analysis. Alignment of multiple 5S sequences from diploid individuals was straightforward, as there were low levels of nucleotide substitution and length variation. Sequence alignment was more problematic, however, for sequences from different diploid genome groups and from different subgenomes of the allopolyploid species. A global alignment among all 99 5S rDNA sequences was accomplished using the PILEUP module of the Wisconsin GCG computer package, version 8.0 (Devereux et al. 1984). After visual inspection of the resulting alignment, minor changes were made in the intergenic spacer region. The alignment used is shown in Fig. 1.

To quantify the diversity of 5S rDNA sequences detected at various levels of organization (e.g., within individuals, within genome groups), we calculated two descriptors: p_n, the proportion of nucleotide sites that are polymorphic and π, nucleotide diversity (Nei, 1987). These calculations were facilitated by using the Molecular Evolutionary Genetics Analysis software (MEGA v. 1.0; Kumar et al. 1993). To test for equivalent patterns of sequence evolution in the 5S gene and spacer regions, we made 2 × 2 contingency tables (McDonald and Kreitman 1991; Kellogg and Appels 1995) for selected taxa, where observed fixed and polymorphic differences (columns) were tabulated for 5S gene and spacer sequences (rows). Because the number of expected fixed differences in the 5S gene was low, two-tailed Fisher exact tests (Sokal and Rohlf 1981) were used to determine significance.

Gene trees of aligned 5S rDNA sequences were generated using maximum parsimony and distance-based methods of phylogenetic reconstruction. Maximum parsimony analysis was performed using PAUP v. 3.1 (Swofford 1990). Several search strategies were employed in an effort to find the most parsimonious trees. In all analyses, characters and character-state transformations were weighted equally. We explored several alternatives for coding of alignment gaps, including treating all gaps as binary, presence/absence characters, coding them as missing data, and excluding gapped positions from the data set prior to analysis. Separate analyses were performed for the complete data set (5S gene + spacer) and for subsets of the data (5S only, spacer only). In all cases, at least five independent heuristic searches were conducted per data set using the random data addition option, in an effort to find shorter "islands" of trees than those recovered from initial searches (Maddison 1991). To evaluate relative levels of support for individual clades, strict consensus trees were generated for all topologies found that were up to five steps longer than the most parsimonious trees ("decay analysis": Bremer 1988; Donoghue et al. 1992).

For distance-based phylogeny estimation, complete and partial data sets were analyzed using MEGA (Kumar et al. 1993). We translated the observed distances between all pairs of sequences to evolutionarily "corrected" (for superimposed substitutions) Kimura two-parameter distances (Kimura 1980) and subjected the resulting distance matrices to neighbor-joining analysis (Saitou and Nei 1987).

Results

5S rDNA Repeats from Gossypium

We sequenced 99 5S rDNAs from 20 taxa representing diploid and allopolyploid *Gossypium* (Fig. 1). In all species, digestion of genomic DNA with *Bam*HI and subsequent probing with 5S rDNA reveals the characteristic ladder of tandemly repeated genes; 5S rDNA organization is conventional in that each repeat within an array consists of 5S RNA genes separated from one another by intergenic spacers (Sastri et al. 1992). Although the 5S gene is nearly invariant in length (121–122 bp; nucleotides 1–122 in Fig. 1), the nontranscribed intergenic spacer is considerably more variable, ranging from 175 to 191 bp (nucleotides 123–316 of the aligned data set).

Alignment of full-length 5S gene sequences within and between the 20 taxa was simple, requiring only a single nucleotide insertion at position 38 for one sequence from *G. robinsonii.* Visual inspection showed that five 5S gene sequences *(A2D1syn3, tomentosum2, rai-*

```
                        10         20         30         40         50         60         70         80         90        100        110
robinsonii11    GGGTGCGATC ATACCAGCAG TAATGCACCG GATCCCA-TC AGAACTCCGG AGTTAAGCGT GCTTGGGTGA GAGTAGTACT TGGATGGGTG ACCTCCTGGG AAGTCCTCGT
robinsonii8     .......... .........C ....A..... .......A.. .........C .......... .......C.. .......... A....A.... .......... ..........
robinsonii10    .......... .........C .....GC... ......G... ........AC ...C...... .T........ .......... A......... .......... ..........
A2D1syn3        .......... .CC------- ---------- -------.-- ---------- ---------- ---------- ---------- ---------- -----..... ..........
A2D1syn5        .......... .........C .......... .......... .........C .......... .......... .......... .......... .......... ..........
A2D1syn6        .......... .........C .......... .......... ........AC .......... ...A...C.. .......... A......... .....T.... ..........
A2D1syn15       ........C. .........C .....T.... .......... ........AC .......... .......C.. .......... A......... .......... ..........
A2D1syn16       .........T .........C .......A.. ......G... ........AC .......... ......A... .......... A......... .......... ..........
arboreum1       .......... .........C ...AC..... ...T...... ..TG....CC .......... ......CC.. ......A... A..G...... .......... ..........
arboreum2       .......... .........C .......... .......... .........C .......... .......C.. .......... .......... .......... ....T.....
arboreum4       .......... .........C .......... .......... ........AC .A........ .......C.. .......... A......... .......... ..........
arboreum5       .......... ........TC .......... .......... ........AC .......... ......CCA. .......... A......C.. .......... ..........
arboreum6       .......... .........C .......... .......... ........AC .......... .......... .......... A......... .......... ..........
herbaceum1      .......... .........C .......... .......... .........C .......... .......C.. .......... A......... .......... ..........
herbaceum2      .......... .........C .......... .......... .........C .......... .......C.. .......... A......... .......... ..........
herbaceum3      .......... .........C .......... .......... .........C .......... .......... .......... A......... .......... ..........
herbaceum4      .......... .........C .......... .......... ........AC .......... .......C.. .......... .......... .......... ..........
herbaceum5      .......... .........C .......... ........A. .........C .......... .......C.. .......... A......... .......... ..........
barbadense1     .......... ...T.....C .......... .......... ........AC .......... .......... .........A A......... .......... ..........
barbadense2     .......... .........C .......... .......... .........C .....G.... .......... .......... .......... .......... ..........
barbadense3     .......G.. .........C .......... .......... .........T .......... .......... .......... .........A .......... ..........
barbadense6     .......... ...T...G.C .......... .......... .........C .......... .......CT. .......... A......... .......... .T........
barbadense7     .......... ...T.....C .....T.... .......... ........AC .......... ....A..C.. .......... .......... .......... ..........
barbadense8     .......... ...T.....C .......... .......... .........C .......... .......C.. .......... A......... .....T.... ..........
barbadense10    .......... ........TC .......... .......... .........A ........A. .......... .......... A......... .......... ..........
tomentoaum1     .......... ...A.....C .......... .......... ........AT ...G...... .......... .........A .......... ....T..... ..........
tomentosum2     ...A.A.--- ---------- ---------. .......... ........AC ...C...... .......C.. ..A....... A......... .......... ..........
tomentosum6     .......... .........C .......... .......... .........C .......... .......... .......... A......... .......... ..........
tomentosum7     .......... .........C .......... .......... .A.......T .......... .......C.. .......... AA........ .......... ..........
tomentosum10    .....T.... .........C .......... .......... .........T .......... .......... .......... A......... .....T.... ..........
tomentosum11    .......... .........C .......... .......... .........T .......... .......... .......... A......... .......... ..........
mustelinum9     T......... .........C .......... .......... .........C .......... .......C.. .......... .......... .....T.... ....TT....
mustelinum15    A......... .........C .......... .......... .......T.C ........A. ...C..TA.. .......... A........A .....A.... ..........
mustelinum16    A......... ....T....C .......... ......G... ........AC ........T. .......... .......... .......... .......A.. ..........
mustelinum18    .......... .........C .......... ......G... ........AC .......... .......... .......... A......... .....T.A.. ..........
mustelinum19    .......... .....G...C .......... .......... ........AC .......... .......... .......... A......... ........C. ..........
hirsutum1       .....A.... .........C .......... .......... ........AC .......... ..A....... .......... .......... .......... ..........
hirsutum5       .......... ..T.A....A .......... .......... ....A....C .......... .......C.. T......... A......... .......... ..........
hirsutum8       ...A..A... .........C .......... .......... .........C .T........ .......C.. .......... .......... .......... ..........
hirsutum9       ....C..... .........C .......... .......... .........C .......... .......C.. .......... A......... .......... ..........
hirsutum11      .......... ...A.....C .......... .......... ........AC .......... .......... .......... A......... .......... ..........
gossypioides1   .......... ...TA....C .......... .........T .........C .......... .......C.. .......... A......... .T........ ..........
gossypioides4   .......... .........C .......... .......... .........C .......... .......C.. .......... AC........ .........A ..........
gossypioides5   .......... .........C .......... .......... .........C .......... .......C.. .......... AC........ .......... ...A....T.
gossypioides6   ......A... .........T ....C..... .......... ........AC T......... .......C.. .........A .......... .......... ..........
schwendemanii5  ........C. .........C .......... .......... .........C .......... .......C.. .......... A......... .......... ..........
schwendemanii8  .......... .........C .......... .......... .........C .A........ .T.....C.. .......... .......... .....T.... ..........
schwendemanii6  .....T.... .........C .......... .......... .........C .......... T......C.. .......... A......... .......... ..........
schwendemanii7  .......... .........C .......... .......... .........C .......... .......C.. ...C...... A......... .......... .........C
aridum4         .......... .........C CCGG.T.... .......... ......A..C .......... .......C.. .......... A......... .......... .........?
aridum5         .......... .........C .......... .......... .........C ........A. .?.....C.. .......... A......... .......... .........A
lobatum4        .......... .........C .......... .......... .........C ........A. .......C.. .......... A......... .......... .........A
lobatum1        .......... .........C .......... .......... .........C .......... .......C.. .......... A......... .......... ..........
lobatum3        .......... ....T....C .......... .......... .........C .......... .......C.. .......... .......... .......... ....A....A
lobatum5        .......... .......?.C ....CAC.-. .......... ........AC .......... ....C..C.. .......... .......... .....T.... ..........
lobatum2        .........T .........C .......... ......G... ........AC .......... .......C.. .......... A......... .....T.... ..........
laxum4          .......... .........C .......... .......... ........AC .......... .......C.. .......... .......... .....T.... ..........
laxum5          .......... .........C .......... .......... ........AC .......... .......... .......... A......... .......... ..........
laxum6          .......... .........C .......... .......... .........C .......... .......C.. .......... .......... .......... ..........
klotzschianum4  .......... ..T......C .......... .......... ........AC .......... .......... .......... A......... .....A.... ...C......
klotzschianum5  .......... .........C .......... .......... ........AC .......... .......C.. .......... A......... .......... ..........
klotzschianum6  .......... ...T.....C .......... .......... .........C .......... .......C.. .......... A......... ..T....... ..........
klotzschianum8  .......... ..T......C .......... .......... ........AC .......... .......C.. .......... .......... .......... ..........
davidsonii5     T......... ....T....C .......... .......... ........AC .......... .......... .......... A......... .......... ..........
davidsonii6     ......A... ..T.?....C .......... .......... .........C .......... .......C.. .........A A......... .......... ..........
trilobum1       .......... .........C .......... .......... .........C .......... .......C.. .......... .......... .......... ..........
trilobum10      .......... .........C ....A..... .......... ........AC .......... .......C.. .........A A......... .......... ..........
trilobum11      .......... ........TC .......... .......... .......T.C .......... .......C.. .......... A......... .......... ....T.....
thurberi3       .......... .........C .......... .....T.... .........C .......... .......C.. .......... A......... .......... ..T.......
thurberi5       ....A..... .........C .......... .......... ........AC .......... .......C.. .......... A....C.... ........-. ..C.......
thurberi7       .......... .........C .......... .......... .........C .......... .......C.. .........A .......... .....T.... .T........
A2D1syn11       .......... .........C .......... .....T.... .........C .......... .......C.. .......... A......... .......... ..........
A2D1syn13       .......... .........C .......... .......... ........AC .......... .......... .........A A......... .......... .......T..
A2D1syn14       .........T .........C .......... .......... ........AC .......... .......... .........A .......... .......... .G........
A2D1syn17       .......... .........C .......... .......... ........AC .......... ..C....... .......... .......... .......... ..........
turneri1        .......... .........C .......... .......... .........C .......... .......C.. .......... A......... .......... ..........
turneri2        .......... .........T .......... .......... ........AC .......... .......... .......... .......... .......... ........T.
turneri3        .....T.... .........C .......... .......... .........C .......... .......C.. .......... A......... .......... ..........
harknessii1     .......... .........C .......... ......G... ........AC .......... ..C....... .......... .......... .......... ..........
harknessii2     .......... ........TC .......... ......G... ........AC .......... ....T..C.. .......... A......... .......... ...AA.....
harknessii3     .......... .........C .....TG... .......... .........C .A........ ....T..C.. .......... A......... .......A.. ..........
harknessii5     .......... .........C .......... .......... ........AC .......... .......C.. .......... G......... .......... ..........
armourianum1    ...C...... .........C ..T....... .......... .........T ........T. .......... .......G.. A......... .......... ...G.....C
armourianum10   .......... ........CC .......... .......... ........AC .......... .......... .......... A......... ......A... ..........
raimondii1      .......... .........C .......... .......... .........T ....C..... .......C.. .......... A......... .......... ..........
raimondii3      .......... .........C .........- -------.-- ---------- ---------- --------.. .......... .......... .......... ..........
raimondii4      .......... .........C .......... .......... .........C ..A....... .......... .......... A......... .......... ..........
raimondii5      .......... .........C .......... ......T... .........C .......... .......C.. .......... A......... .......... ..........
raimondii6      ---------- ---------- ---------. .......... ........AC .......... .......C.. .......... A......... .......... ..........
raimondii7      .......... .........C .......... ......G... .........C .......... .......C.. .......... A.A....... .......... ..........
mustelinum6     .......... .........A .......... .......... ........AC .......... ....C..... .........A .......... .......A.. ........C.
mustelinum7     .......... .........C ....A..... .......... .........T .......... .......... .......... A......... ........A. ..........
mustelinum13    .......... .........C .......... .......... ........AC .......... .......... .........A A......... ........A. ..........
mustelinum14    .......... .........C .......... ......G... ........AC .......... .......... .......... A......... ........T. ..........
mustelinum17    .......... .........C .......... .......... ......A..C ..G...A... .......C.. .......... AC........ .......... T.........
hirsutum6       .......... .........C ....?.?..- .C.TTT.... T..TAGAAT. C.C...AA.. .A......C. .-TC.TAC.A GCAC.AATGC ....T..... ..........
hirsutum7       ---------- ---------- ---------. .......... .........C .......... .......... .........A A......... .......A.. ..........
hirsutum10      .......... .........C .........- .CCTTT...A TA.TAGAAT. C.CA.GAAT. .A.G..TG.. .......... .......... .....T.... ....T.....
hirsutum12      .......... ....T....C .......... .......... ........AC .......... .......C.. .......... .......... ....T..... ..........
```

Fig. 1. Aligned nucleotide sequences of 99 cloned 5S rDNA repeats from *Gossypium. Periods* represent residues identical to those of the reference sequence from *G. robinsonii; dashes* indicate alignment gaps; and *question marks* denote missing information. The first 122 nucleotides correspond to the 5S gene; the intergenic spacer encompasses nucleotides 123–316.

mondii3, raimondii6, and *hirsutum7)* were considerably shorter than the expected length. We feel that these truncated sequences are deletion artifacts of cloning rather than actual sequences since M13 can delete insert DNA (Sambrook et al. 1989). In contrast to the 5S gene, alignment of the spacer region was not trivial, due to higher sequence divergence between taxa from different genomic groups. Indels were introduced in a region of relatively low conservation in the 5' half of the spacer region (nucleotides 135–170 and 182–205, Fig. 1), two of which are genome-specific. Indel 1 is 7 bp in length (nucleotides 183–189) and is interpreted as either a deletion in

```
                       120        130        140        150        160        170        180        190        200        210        220
robinsonii11    GTTGTTCCCC TCCCATTTTA TTTTATTCCG TATGAATTTT CTTTTCCTCT TT-AAAAA-- CATCGTTTAA CTAGTGGGCA ATTGGGTGAG TCGTCACTTG CGAAATAAAT
robinsonii8     .....A.... ...T...... .......... .........C A......... .......... .......... .......... .......... ...C...... ..........
robinsonii10    T....A..-- .......... .......... .........C .......... .......... .......... .......... ....A..... .......... ..........
A2D1syn3        ....CA.... ...A...... .A.CTCATTA G.-TTT.... T-C.ATT.T. ..T..T.TCT .......... ..-------. ...A------ --..T.T..A ..G.......
A2D1syn5        ....CA.... ...A...... .A.CTCATTA G.ATTT.... T-C.ATT.T. ..T..T.TCT .......... ..-------. ...A------ --..T.T..A ..G.......
A2D1syn6        ....CA.... ...A...... .A.CTCATTA G.-TTT.... T-C.ATT.T. ..T..T.TCT .......... ..-------. ...A------ --..T.T..A ..G.......
A2D1syn15       ....CA.... ...A...... .A.CTCATTA G.-TT..... T-C.ATT.T. ..T..T.TCT .......... ..-------. ...A------ --.AT.T..A ..G.......
A2D1syn16       ....CA.... ...A...... .A.CTCATTA G.-TTT.... T-C.ATT.T. ..T..T.TCT .......... ..-------. ...A------ --C.T.T..A .A........
arboreum1       ....CA.... ...A....G. .A.CTCATTA G.-TTT.... T-C.AAT.T. ..T..T.TCC .....C.... ..-------. ...A------ --..T.T... ..G.......
arboreum2       ....CA.... .......... .A.CTCATTA G.-TT..... T-C.ATT.T. ..T..T.TCT .......... ..-------. ...A------ --..T.T..A ..G.......
arboreum4       ....CA.... C..A...... .A.CTCATTA G.-TTT.... T-C.ATT.T. ..T..T.TCT .......... ..-------. ...A------ --..T.T..A ..G.......
arboreum5       ....CA.... ...A...... .A.CTCATTA G.-TTT.... T-C.ATT.T. ..T..T.TCT .......... ..-------. ...A------ --..T.T..A ..G.......
arboreum6       ....CA.... ...A...... .A.CTCATTA G.--TT.... T.C.ATT.T. ..A-.T.TCT .....C.... ..-------. ...A------ --..T.T..A ..G.......
herbaceum1      ....CA.... ...A...... .A.CTCATTA A.-TTT.... T-C.ATT.T. ..TT.T.TCT ......C... ..-------. ...A------ --..T.T..A T.G.......
herbaceum2      ....CA.... ...A...... .A.CTCATTA A.-TTT.... T-C.ATT.T. ..TT.T.TCT ......C... ..-------. ...A------ --..T.T..A T.G.......
herbaceum3      ....CA.... ...A...... .A.CTCATTA A.-TTT.... --C.ATT.T. ..TT.T.TCT ......C... ..-------. ..AA------ --..T.T..A ..G.......
herbaceum4      ....CA.... ...A...... .A.CTCATTA A.-TTT.... --C.ATT.T. ..TT.T.TCT ......C... ..-------. ...A------ --..T.T..A ..G.......
herbaceum5      ....CA.... ...A...... .A.CTCATTA A.-TTT.... --C.ATT.T. ..TT.T.TCT ......C... ..-------. ...A------ --..T.T..A T.G.......
barbadense1     ....CA.... ...A...... .A.CTCATTA A.---T.... T.C.ATT.T. ..T-.T.TCT A......... ..-------. ...A------ --..T.T..A ..G.......
barbadense2     ....CA.... ...A...... .A.CTCATTA A.--TT.... T.C.A-T.T. ..T-.T.TCT .......... ..-------. ...A------ --..T.T..A ..G.......
barbadense3     ....CA.... ...A...... .A.CTCATTA A.A--T.... T.C.ATT.T. ..T-.T.TCT .......... ..-------. ...A------ --..T.T..A ..G.......
barbadense6     ....CA.... ...A...... .A.CTCATTA A.--TT.... T.C.A-T.T. ..T-.T.TCT A......... ..-------. ...A------ --..T.T..A ..G.......
barbadense7     ....CA.... ...A...... .A.CTCATTA A.--TT.... T.C.ATT.T. ..T-.T.TCT ..A....... ..-------. ...A------ --..T.T..A ..G.......
barbadense8     ....CA.... ...A...... .A.CTCATTA A.--TT.... G.C.ATT.T. ..T-.T.TCT .....CA... ..-------. ...A------ --..T.TC.A ..G.......
barbadense10    ...TCA.... ...A...... .A.CTCATTA A.--TT.... T.C.ATT.T. ..T-.T.TCT ......A... ..-------. ...A------ --..T.T..A ..G.......
tomentosum1     ....CA.... ...A...... .A.CTCATTA A.--TT.... T-C.A-T.T. ..T-.T.TCT .......... ..-------. ...A------ --..T.T..A ..G.......
tomentosum2     ...ACA...T ...A...... .A.CTCATTA A.---T.... T.C.A--.T. ..TT.T.TCT .G........ ..-------. ...A------ --..T.T..A ..G.......
tomentosum6     .C..CA...T ...A...... AA.CTCATTA A.--TT.... T.C.A-T.T. ..T-.T.TCT .......... ..-------. ...A------ --..T.T..A ..G.......
tomentosum7     ....CA.... ...A...... .A.CTCATTA A.--TT...C T.C.ATT... ..T-.T.TCT .....CA... ..-------. ...A------ --..T.T..A ..G.......
tomentosum10    ....C..... A..A...... .A.CTCATTA A.--TT.... T.C.A-T.T. ..T-.T.TCT .......... ..-------. ...A------ --..T.T..A ..G.......
tomentosum11    ....CA.... ...A...... .A.CTCATTA A.---T.... T.C.ATT.T. ..T-.T.TCT .......... ..-------. ...A------ --..T.T..A ..G.G.....
mustelinum9     ....CA.... ...A...... .A.CTCATTA A.--TT.... T.C.GTT.T. ..T-.T.TCT .......... ..-------. ...A------ --..T.T..A ..........
mustelinum15    ....CA.... ...A.....T .A.CT.ATAA A.--TT.... T.C.A-T.T. ..T-.T.TCT ......A... ..-------. ...A------ --..T.T..A ..G.......
mustelinum16    A...CA...T ...A...... .A.CTCATTA A.---T.... T.C.ATT.T. ..T-.T.TCA ......A... ..-------. ...A------ --..T.T..A .AG.......
mustelinum18    ....CA.... ...A...... .A.CTCATTA A.--TT.... T.C.A-T.T. ..T-.T.TCT .......... ..-------. ...A------ --..T.T..A A.G.......
mustelinum19    ....CA.... ...A...... .A.CTCATAA A---TT.... T.C.GTT.T. ..T-.T.TCT .......... ..-------. ...A------ --..T.T..A A.G.......
hirsutum1       ....CA..AT ?..A...... .A.CTCATTT A.---?.... T.C.A-T.TG ..TT.T.TCT ......A... ..-------. ...A------ --..T.T..A ..G.......
hirsutum5       ....CA.... ...A...... .A.CTCATTA A.--TT.... T.C.ATT.T. ..T-.T.TCT ......A... ..-------. ...A------ --..T.T..A ..G.......
hirsutum8       ....CA.... ...A...... .A.CTCATTA A.--TT.... T.C.ATT.T. ..T-.C.TCT ......A... ..-------. ...A------ --..T.T... ..G.......
hirsutum9       ....CA.... ...A...... .A.CT.ATTA A.--TT.... T.C.ATT.T. ..T-.T.TCT ......A... ..-------. ...A------ --..T.T..A ..G.......
hirsutum11      ...TCA...T C..A...... .A.CTCATTA A.--TT.... T.C.ATT.T. ..T-.T.TCT ......A... ..-------. ...A------ --..T.T... .AG.......
gossypioides1   ....CA.... ....T...GG ...C..AAAA .---TC.... T.G..-GCA. .ATC-T.-AA .G..ACG... ..-------. ...A-..T.T .G.CGTT..A ..........
gossypioides4   ....CA.... ....T....G ...C..AAAA .---TT.... TAG..-GCA. .ATC-T.-AA .G..ACG... ..-------. ...A-C.T.T .G?CGTT... T.....G.C.
gossypioides5   ....CA...- -----....T ..A...AAA. .T.TTTG... TAG.A-GCA. ..TC-T.-AA ..C...G... ..-------. ...A--.TGT .TC.ATT... ..........
gossypioides6   ....AA...G ....T....G ...C..ATAA .---T-.... TAG..-GCA. .ATC-T.-AA .G..A.G... ..-------. ...A-..T.T .?CCGTT... ..........
schwendemanii5  ....CG.... .......... ...CG.A.AA .G-TTT.... TAG..-G.A. ..GC-TC-CG AG.G..G... ..-------. ...C-..T.. .T.CGTT... ..........
schwendemanii8  ....CA.... .......... ...CG.ATAA .G-TTT.... TAG..-GCA. ..GC-TC-CG .G.G..G... ..-------. ...C-..T.. .T.CGTT..A ..........
schwendemanii6  ....CA.... .......... ...C..ATAA .G-TTT.... TAG..-GCA. ..CC-TC-CA .G.G..G..G ..-------. ...C-..T.. .T.CGTTG.C ..........
schwendemanii7  ....CA.... ..TT...... ...C..ATAA .G-TTT.C.. TAG..-GCA. ..GC-TC-CG .G.G..G... ..-------. ...C-..T.. ...CGTT... ..........
aridum4         ....CG.... ..TT...... ...C..ATAA .G-TTT.... TAG..-GCA. ..GC-TC-CG .G.G..G... ..-------. ...C-..T.. .T..ATT... ..........
aridum5         ....CA.... ....T..... ...C..ATAA .G-TTT.... TA...-GCA. ..TC-TC-CG .G.GA.G... .A-------. ...C-..T.. .T.CGTT... ..........
lobatum4        ....CA.... ....T..... ...C..ATAA .G-TTT.... TA...-GCA. ..TC-TC-CG .G.GA.G... .A-------. ...C-..T.. .T.CGTT... ..........
lobatum1        ....CA.... ....T..... ...C..ATAA .G-TTT.... TAG..-GCA. ..TC-TC-CT .G.G..G... .A-------. ...C-..T.. .T.CGTT... ..........
lobatum3        ....CA.... ....T..... ...C..ATAA .G-TTT.... TA...-GCA. ..TC-TC-CG .G.G..G... .A-------. ...C-..T.. .T.CGTT.C. ..........
lobatum5        ....CA.... ....T..... ...C..ATAA .G-TTT.... TA...-GCA. ..TC-.C-CG .G.G..G... .A-------. ...C-..T.. .T.CGTT... ..........
lobatum2        ....CA.... ....T..... ...C..ATAA .G--TT.... TA...-GCA. ..TC-TC-CG .G.G..G... .A-------. ...C-..T.. .T.CGTT... ..........
laxum4          ....CA.... .......... ...C..ATAT .G.TTT.... TCG..-GCA. ..TC-CC-CA .G.G..G... .A-------. ...C-..T.C ...CGTT..T ..........
laxum5          ....CA.... .......... ...C..ATAT .G.TTT.... TAG..-G.A. ..TC-CC-CA .G.G..G... .A-------. ...C-..T.C ...CGTT..T ..........
laxum6          ....CA.... ...T...... ...C..ATA. ..-TTT.... TAG..-GCA. ..TC-CC-CA .G.G..G... .A-------. ...C-..T.. .T.CGTT... ......T...
klotzschianum4  ....C..... ..A....... .G.C..ATAA ..-TTT..-. TAG..-GCA. ..TC-TC-CA .G....G... ..-------C ...C-..T.T .T.CGTT... .......G..
klotzschianum5  ....CA.... .......... ...C..ATAA ..-TTT..-. TAG..-GCA. ..TC-TC-CA .G....G... ..-------C ...C-..T.T .T.CGTT... ..........
klotzschianum6  ....CA.... .......... .G.C..ATAA ..-TTT..-. TAG..-GAA. ..TC-TC-CA .G....G... ..-------C ...C-..T.T .T.CGTT... ..........
klotzschianum8  ....CA.... .......... .G.C..ATAA ..-TTT..-. TAG..-GCA. ..TC-TC-CA .G....G... ..-------C ...C-..T.T .T.CGT...A ..........
davidsonii5     ....CA.... ..T....... .G.C..ATAA ..-TTT.--. TAG..-ACA. ..TC-TC-CA .G.T..G... ..-------C .C.C-..T.T .T.CGTT... ..........
davidsonii6     ....CA.... A......... .G.C..ATAA ..-TTT.... TAG..-GCA. ..TC-TC-CA .C.G..G... ..-------C ...C-..T.T .TTCGTT... ..........
trilobum1       ....CA.... .......... ...C..ATAA ..-TTT..-. TCG..-GCA. ..TC-TC-CA .G.T..G... .G-------C ...C-CCT.T .T.CGTT..C ..........
trilobum10      ....CA.... .......... ...C..ATAA ..-TTT..-. TCG..-GCA. ..TC-TC-CA .G.G..G... .G-------C ...C-CCT.T .T.CGTT..C ..........
trilobum11      ....CA.... .......... ...C..ATAA ..-TTT..-. TCG..-GCA. ..TC-TC-CA .G.G..G... ..-------C ...C-C.T.T .TACGTT... ..T.......
thurberi3       ....CG...T .......... ...C..ATAA ..-TTT.--. TCG..-GCA. ..TC-TC-CA .G.G.AG... ..-------C ...C-C.T.T .T.CGTT..C ..........
thurberi5       ....CA...T .......... ...C..ATAA ..-ATT.... TAG..-GCA. ..TC-TC-CA .G.G..G... .C-------C ...C-C.TGT .T..GTT... ..........
thurberi7       ....CA.... ...T.C.... ...C..ATAA ..-TTT..-. TCG..-GCA. ..TC-TC-CA .G.G..G... .G-------C ...C-C.T.T .T.CGTT..C ..........
A2D1syn11       ....CA.... ........A. ...C..ATAA ..-TTT..-. .CG..-GCT. ..TC-TC-CA .G.G..G... .G-------C ...C-C.T.T .T.CGTT..C ..........
A2D1syn13       ....CA.... .......... ...C..ATAA ..-TTT..-. TCG..-GCA. ..TC-TC-CA .G.G..G... .G-------C ...C-C.T.T .TTCGTT..C ..........
A2D1syn14       ....CA.... .......... ...C..ATAA ..-TTT..-. TCG..-GCA. ..TC-TC-CA .G.G..G... ..-------C ...C-C.T.T .T.CGTT... G.........
A2D1syn17       ....CA.... .......... ...C..ATAA ..-TTT..-. TCG..-GCA. ..TC-TC-CA .G.G..G... .G-------C ...C-C.T.T .T.CGTT..C ..........
turneri1        ....CA.... .......... ...CG.ATAA .G-TTT..G. TAG..-GCA. ..TC-TC-CA .G.G..G... ..-------C ...C-..T.C .T.CGTT..A ..........
turneri2        ....CA.... ..A....... ...C..ATAA AG-TTT..G. TAG..-GCA. ..TC-TC-CA .G.G..G... ..-------C ...C-..T.T .T.CGTT... ..........
turneri3        ....CA.... .......... ...CG.ATAA .G-TTT..G. TAG..-GCA. G.TC-TC-CA .G.G..G... ..-------C ...C-..T.C .T.CGTT..C ..........
harknessii1     ....CA.... .......... ...CG.ATAA .G-TTT.CA. TAG..-GCA. ..TC-TC-CA .G.G..GC.. ..-------C ...A-T.T.T .T.CGTT... ..........
harknessii2     ....CA.... ...A...... ...CC.ATAA .G-TTT..G. TAG..-GCA. ..TC-TC-CA .G.G..G... ..-------C ...C-..T.T .T.CGTT... ..........
harknessii3     ....CA.... .......... ...CG.ATAA .G-TTT..G. TAG..-GCA. ..TC-TC-CA .G.G..G... ..-------C ...C-..T.T .T.CGTT... ..........
harknessii5     ....CA.... .....A.... ...CG.ATAA .G-TTT..G. TAG..-GCA. ..TC-TC-CA .G.G..G... ..-------C ...C-..T.C .T.CGTT... ..........
armourianum1    ....CA.... .......... ...CG.ATAA .G-TTT..G. TAGG.-GCA. ..TC-TC-CA .G.T..G.C. ..-------C ...C-..T.T .T.?GTT... ?.........
armourianum10   ....CA.... .......... ...C..ATAA .G-TTT..G. TAG..-GCA. ..TC-TC-CA .G.G..G... ..-------C ...C-..T.T .T.CGTT... ..........
raimondii1      ....CA.... .......... ...C..AGAA .T-TTT.... TAGG.-GCA. ..TC-TC-GA .G.G..G.C. ..-------C ...C-.GT.T .TTCGTTAG. T.........
raimondii3      ....CA.... .......... ...C..ATAA .G-TTT.... TAG..-GCA. ..TC-TC-GA .G.G..G... ..-------C ...C-..T.T .T.CGTT... T........C
raimondii4      ....CAT... ..T....... ...C..ATAA .T-TTT.--. TAG..-GCA. ..TC-TC-GA .G.G..G... ..-------C .A.C-..T.T .T.CGTT..C G.........
raimondii5      ....CA.... .......... ...C..ATAA .G-TTT.... TAG..-GCA. ..TC-TC-GA .G.G..G..G ..-------C ...C-..T.T .T.CGTT... T........C
raimondii6      ....CA.... .......... ...C..ATAA .T-TTT.--C TAG..-GCA. ..TC-TC-GA .G.G..G... ..-------C .A.C-..T.T .T.CGTT... T........C
raimondii7      ....CA.... .......... ...C..ATAA .G-TTT.... TAG..-GCA. ..TC-TC-GA .G.G..G... ..-------C ...C-..T.T .T.CGTT... T.........
mustelinum6     ....CA...T .........G ...C..ATAA .G-TTT.... TAG..-GCG. ..T.-TC-CA .G.G..G... ..-------C ...C-..T.T .T.CGTT... TA........
mustelinum7     ....CA.... .......... ...C..ATAA .G-TTT.--. T.G..-GCG. ..TC-TC-CA .G.G..G... ..-------C ...C-..T.T .T.CGTT... T.........
mustelinum13    ....CA.A.. .......... ...C..ATAA .G-TTC.... TAG..-GCG. ..TC-TC-CA .G.G..G... ..-------C ...C-..TCT .T.GGTT... TA........
mustelinum14    ....CA.G.. .......... ...C..ATAA .G-TT-.... TAG..-ACG. ..TC-.C-CA .G.G..G... ..-------C ...C-..T.T .T.CGTT... T.........
mustelinum17    ....CA.... .......... ...C..AGAA .G-TTT.... T.G..-GCG. ..TC-TC-CA .G.G..G... ..-------C ...C-..T.T .T..GTT... T.........
hirsutum6       C...CA.... .........G ....CA.ATA ATGTTT..G. TAG..-GCG. ..TC-TC-CA .G.G..G... ..-------C ...C-..T.T .T.CGTT... T.........
hirsutum7       ....CA.... ...TG..... ...C..ATAA .G.TTT.... TAG..-TCG. ..TC-TC-CA TG.GT.G... ..-------C ...C-A.T.T .TTCGTT... T.........
hirsutum10      ....CA.... .......... ...C..ATAA C.--TT.... TAG..-GCA. ..GC-TC-CA .G.G..G... ..-------C ...C-..T.T .T.CGTT... .......T..
hirsutum12      ....CA.... ?......... ...C..ATAA .G-TTT.... TAG..-ACG. ..TC-TC-CA .G.G..G... ..-------C ...C-..T.T .T.CGTT... T.........
```

Fig. 1. Continued.

the 5S rDNA of the common ancestor of A- and D-genome diploid species or as an insertion in the *G. robinsonii* lineage. Indel 2 is 8 bp in length (nucleotides 195–202) and is unique to sequences from A-genome diploids, implying that a deletion occurred in 5S rDNA of the common ancestor of these species. These diagnostic indels, combined with phylogenetic analyses (detailed below), allowed us to infer the subgenomic origin of each allopolyploid sequence. Clones from the A-subgenome were recovered from all allopolyploid species studied, whereas clones originating from the D-subgenome were detected only in *G. mustelinum, G. hirsutum,* and the synthetic polyploid $2(A_2D_1)$. As with the 5S gene sequences, apparent M13 deletion artifacts

```
                       230        240        250        260        270        280        290        300        310    316
robinsonii1     AAATTGAAAC GTTATTTTGT CAAATTTACA TTGAAATTCG AGGCGGAGGT GCGATAAGGG GAAGCCTTTA TATAAATAAA TTGCGCAAGA GTTAAC
robinsonii8     .....A.... ......CG.. .G........ .......... .....C.... .T.....T.. .C........ .......... .......... ......
robinsonii10    .....A.... .......G.. .G........ .......... .....CT... .......... .C........ .......... .......... T.....
A2D1syn3        C.C.C..... .......G.. .GT.....A. ..C....... ....TC.A.C .....C.... .T........ ........G. A.A....... A..G..
A2D1syn5        C.C.C..... .......G.. .GT.....A. ..C....... ..A.TC.A.C .....C.... TT........ ........G. A......... A..G..
A2D1syn6        C.C.C..... .......G.. .GT.....A. ..C....... ....TC.AAC .....C.... .T........ ........G. A......... A..G..
A2D1syn15       C.C.C..... .......G.. .GT.....A. ..C....G.. ....TC.A.C .....C.... .T........ ........G. A.A....... A..G..
A2D1syn16       C.C.C.C... .C.....G.. .GT.....A. ..A...A... ....TC.A.. .....C.... .T.C...... ........G. A......... A..G..
arboreum1       C.C.C..... .......G.. .GT..C..A. ..C....... ....TC.A.C .....C.... .T........ ........G. A......... A..G..
arboreum2       C.C.C..... .......G.. .GT..G..A. ..C....... ..C.TCCA.C .....C.... .T........ ........G. A.A....... A..GCA
arboreum4       T.C.C..... .......G.. .GT..C..A. ..C....... ....TC.A.C .....T.... .T...A.... ........G. A.A....... A..G.T
arboreum5       C.C.C..... .......G.. .GT.....A. ..C....... ....TC.A.C .....C.... AT........ ........G. A......... A..GC.
arboreum6       C.CAC....T .......G.. .GT.....A. ..C....... .T..TC.A.C .....C.... .T..?..... .G......G. A.A....... A..G..
herbaceum1      C.C.C..... .......G.. .GT.....A. ..C....... ..T.TC.A.C .....C.... .T........ ........G. A...A..... A..G..
herbaceum2      C.C.C..... ....C..G.. .GT.....A. ..C....... ....TC.A.C .....C.... .T........ ........G. A......... A..G..
herbaceum3      C.C.C..... .......A.. .GT.....A. ..C...C... ....TC.A.C .....C.... .T........ ........G. A..G...... A..G..
herbaceum4      C.C.C..... .......A.. .GT.....A. ..C....... ....TC.A.C .....C.... .T........ ........G. A......... A..G..
herbaceum5      C.C.C..... .......G.. .GT.....A. ..C....... ....TC.A.C .....C.... .T........ ........G. A......... A..G..
barbadense1     C.C.C..... .......G.. .GT.....A. ..C....... ....TC.ACC .....C.... .T........ ........G. A......... A..G..
barbadense2     C.C.C..... .......G.. .GT.....A. ..C....... ....TC.A.C .....C.... .T........ ........G. A......... A..T.T
barbadense3     C.C.C..... .......G.. .GT.....A. ..C....... ....TC.ACC .....C...C .T........ ........G. A.......A. T..G..
barbadense6     CTC.C..... .......G.. .GT.....A. ..A....... ....TC.A.C A....C.... .T........ ........G. A......... A..G..
barbadense7     C.C.C..... .......G.. .GT.....A. ..C....... ....TC.A.C .....C.... .T........ ........G. A......... T..G..
barbadense8     C.C.C..... .......G.. .GT.....A. ..C......A ....TC.ATC .....C.... TT........ ........G. A......... A..G..
barbadense10    C.C.C..... .......G.. .GT..A..A. ..C....... ....TC.ACC .....C...C TT........ ........G. A......... A..G..
tomentosum1     CTC.C..... .......G.. .GT.....A. ..T....... ....TC.A.C .....C.... .T........ ........G. A......... A..G.T
tomentosum2     C.C.C..... .......G.. .GT.....A. ..A......A ....TC.ATC .......... .C........ ...G....G. .......... ...G..
tomentosum6     C.C.C..... .......G.. .GT.....A. ..T....... ....TC.A.C .....C.... .T........ ........G. A......... A..G.T
tomentosum7     C.C.C..... .......G.. ..T.....A. ..C....... ....TC.A.C .....C.... .T..G..... ........G. A......... A..T.T
tomentosum10    C.C.C..... .......G.. .GT.....A. ..T....... ....TC.A.C .....C.... .T........ ........G. A......... A..G.T
tomentosum11    C.C.C..... .......G.. .GT.....A. ..C......A ....TC.ATC .....C.... .T........ ........G. A......... A..G..
mustelinum9     C.C.C.G... .......G.. .GT.....A. ..C....... ....TC.A.C .....C.... .T........ ........G. A......... A..G..
mustelinum15    C.C.C..... .......G.. .GT.....A. ..C....A.. ....TC.A.C .....C.... .T......G. ........G. A......... A..G..
mustelinum16    C.C.C..... .......G.. .GT.....A. ..C....... ....TC.A.C .....C.... .T........ .......... A......... A..G..
mustelinum18    C.C.C..... A......G.. .GT.....A. .......... ....TC.A.C .....C.... .T........ ......T.G. A......... A..G..
mustelinum19    C.C.C..... A......G.. .GT.....A. .......A.. ....TC...C .....C.... .T........ ......T.G. A......... A..G..
hirsutum1       C.CAC..... .......GT. .GT.....A. ..C...C.T. ..?.-C.A.C ....?T.... .?.A.A.... C.......GG A...?.?... A.GG..
hirsutum5       C.G.C..... A......G.A .GC.....A. ..C....... ....TC.A.C .....C.... .T........ ........G. A......... A..G.T
hirsutum8       C.C.C..... .......G.. .GT.....A. ..C....... ....TC.A.C .....C.... .T........ ........G. A......... A.....
hirsutum9       C.C.C..... .......G.. .GT.....A. ..C....... ....TC.A.C .....C.... .T........ ........G. A.T....... A..G.T
hirsutum11      C.C.C..... .......G.. TGT.....A. ..C....... ....TC.A.C .....C.... .T........ ........G. A...?..... A..G.T
gossypioides1   C.T.C..... .C.C...G.. .G.....G.. ..C....... .....C.A.. .T........ .C..AA.... ........G. ..A....G.. .A.G.A
gossypioides4   C.T.C....A .......C.. .G.....G.. ..C......T .....C.A.. .........C .C........ ........G. .......... ...G..
gossypioides5   C.T....... ........C. .G....G... A.C..G.C.. ...T.C.A.. .G......CA .C........ C.......G. C......... ...G..
gossypioides6   C.T.C..... TC.....C.. .G.....GT. ..C....... .....C.AA. .......... .C........ ........G. .......... .A.G..
schwendemanii5  ..T.AA.... .....A.G.. .G.....G.. .CC....... ...T.C.A.. .......... .C........ ...G....G. .......... ...G..
schwendemanii8  ..T.AA.... .....A.G.. .G.....G.T ..C....... .....C.A.. .......... .C........ ...G....G. .......... ...G..
schwendemanii6  ..T.AA.... ..A..A.G.. .G.....G.. ..C....... ....AC.A.. .......... .C........ ...T....G. .......... ...G..
schwendemanii7  ..T.AA...G .....A...C .G.....G.. ..C....... C....C.A.. .......... .C........ ...G....G. .......... ...G..
aridum4         ..T.AA.... .....A.G.C .G.....G.. .CC....... .....C.... .......... .?.A...... ...G....G. .......... ...G..
aridum5         ..T.AA.... .....G..CC .G.....G.. ..C....... .....C.A.. .......... .T........ ...G....G. ........A. ...G.T
lobatum4        ..T.AA.... .....G..CC .G.....G.. ..C....... .....C.A.. .......... .T........ ...G....G. ........A. ...G.T
lobatum1        ..T.AA.... .....C...C .......G.. ..C....... .....C.A.. .......... .T........ ...G....G. .......... ...G..
lobatum3        ..TGAA.... .....C...C .G.....G.. ..C....... .....C.A.. .......... .T........ ...G....G. .......... ...G..
lobatum5        ..T.AA.... .....G...C .G.....G.. ..C....... .....C.A.. .......... .C........ ...G....G. .......... ...G..
lobatum2        ..T.AA.... .....?...C .......G.. ..C....... .....C.A.. .......... .T...T.... ...G....G. .......... ...G..
laxum4          ..T.GA.... .....A.G.C TG.....G.. A.C....... .....C.A.. A......... .C-....... ...G....G. C......... .A.G..
laxum5          ..T.GA.... .....A.G.C TG.....G.. A.C.....T. .....C.A.. .......... .C........ ...G....G. C......G.. ...G..
laxum6          ..T.AA.... .....A.G.C TG.....G.. A.C....... .....C.A.. .....T.... .C........ ...G....G. C......... ...G..
klotzschianum4  ..T....C.. .A.....G.. .G.....G.. ..C....... .....C.A.. .......... .C........ ..CG....G. .......... ...G..
klotzschianum5  ..T....... .A.....G.A .G....CT.. ..C.....T. .....C.A.. .......... .C......A. ...G....G. .......... ...G..
klotzschianum6  ..T....?.. .A.....G.. .G....CT.. ..C....... .....C.A.. .......... .C........ ...G....G. .......... ...G..
klotzschianum8  ..T....... .A.....G.. .G.....G.. ..C....... .....C.A.. .......... .C........ ..CG....G. .......... ...G..
davidsonii5     ..T....... .A.....G.. .G....CT.. ..C....... .....C.A.. .......... .C........ ...T....G. .......... ...G..
davidsonii6     ..T..C.... .A.....G.A .G....CT.. ..C....... ...T.C.A.. .......... .C........ ..CG....GT .......... ...G..
trilobum1       ..T....... .A.....C.. .C.....G.. ..C....... .....C.A.. .......... .C........ ...G....G. .......... ...G..
trilobum10      ..T....... .A.....G.. .G.....G.. ..C....... .A...C.A.. .......... .C........ ...G....G. .......... ...G..
trilobum11      ..T..C.... .A........ .G.....G.. ..A....... .....T.A.. .......... .C........ ...G....G. .A.T...... ...G..
thurberi3       ..T....... .A.....G.A .G.....G.. ..A.....T. .....C.AT. A......... .C........ ...G....T. .......... ...G..
thurberi5       ..T....... CA.....G.. .G.....G.. ..C....... ....TC.A.. .......... TC........ ...T....G. .......... ..AG..
thurberi7       ..T.CT.... .A.....G.. .G.....G.. A.C....... .....C.A.G .......... .C........ ...G....G. .......... ...G..
A2D1syn11       ..T....... .A.....G.. .G.....G.. ..C....... .....C.A.. .......... .C........ ...G....G. ......T... A..G..
A2D1syn13       ..T....... .A.....G.. .G.....G.. ..C....... .....C.A.. .......... .C........ ...G....G. .......... ...G..
A2D1syn14       ..T.C..... .A.....G.. .G.....G.. ..C....... .....C.A.. .......... .C........ ...G....G. .......... ...G..
A2D1syn17       ..T.C....T .A.....G.. .G.....G.. ..C....... .....C.A.. .......... .C........ ...G....G. .......... ...G..
turneri1        ..T....... .A.....G.. .G.....G.. C.C....... .....C.A.. .......... .C........ ...G....G. .......... ...G..
turneri2        ..T....... .A.....G.. .G.....G.. ..C....... .....C.A.. ...G...... .C........ ...G....G. .......... ...G..
turneri3        ..T....... .A.....G.. .G.....G.. ..C....... .....C.A.. .......... .C........ ...G....G. .......... ...G..
harknessii1     ..T....... .A.....G.. .G.....G.. ..C....... .....C.A.. .......... .C........ ...G...... .......... ...GG.
harknessii2     ..T....... .A.....G.. .G.....G.. ..C....... ....AC.A.. .......... .C........ ...G....G. .......... ...G..
harknessii3     ..T..A.... .A.....G.. .G.....G.. ..C....... .....C.A.. .......... .C........ C..G....G. .......... ...G..
harknessii5     ..T....... .A.....G.. .G.....G.. ..C....... .....C.A.. .......... .C........ ...G....G. .......... A..G..
armourianum1    ..T....... .A.....G.. .G.....G.. ..C....... ...?.C.A.. .....G.... .C........ ...G....G. .......... ...G..
armourianum10   ..T....... .A.....G.. .G.....G.. ..C....... .....C.A.. ....C..... .C........ ...G....G. .......... ...G..
raimondii1      ..TC..G... .A....AG.. .G.....G.. ..T.T..... .....C.A.. .......... .T........ ...G....G. ...T...... ...G..
raimondii3      ..T....... .A.....G.. .G.....G.. ..T.....T. .....C.AC. .......... .C........ ...G....G. ...T...... ...G..
raimondii4      ..C....... .A.....G.. .G..G..G.. ..T....... .....C.AC. .......... .C........ ...G....G. .......... .C.G..
raimondii5      ..T....... .A.....G.. .G.....G.. A.T....... .....C.AC. .......... .C........ ...G....G. .......... ...G..
raimondii6      ..T....... .A.....G.. .......G.. ..T....... .....C.A.. .......... .T........ ...G....G. .....G.--- ------
raimondii7      ..T...G... .A.....G.. .G.....GT. ..C..T.... .....C.A.. A.......C. AC.A...... ...G....G. .......... ......
mustelinum6     ..T...G... .A.....G.. AG.....T.. ..T....... .....C.A.. C......... .C.A..C... ...G....G. .......... ...G..
mustelinum7     ..T...G... .A.....G.. .G.....GTT ..A....... .....T.A.. ........A. .G........ ...G....G. ..C....... ...G..
mustelinum13    ..T...G... .A.....G.. .G.....G.. ..T....... .....C.A.. ........C. .CT....... ...G....G. .......CT. ...G..
mustelinum14    ..T...G... .A.....G.. .G........ ..T....... .....C.A.. .......... .C........ ...G....G. .......... ...G..
mustelinum17    ..C..CG... .A.....G.. .G.....G.. ..T....... .....C.A.. C......... .C.T...... ...G....G. .......... ...G..
hirsutum6       ..T...G... TA.....G.. .G.....GT. ..A....... .....C.A.. .T........ .C........ ...G...CG. ....C..... ...G..
hirsutum7       ..T...C... .A.....G.. .G.....G.. ..T....... .....C.A.. .......... .C........ ...G....GT ...G------ ------
hirsutum10      ..T...G..T .A.....G.. .G.....G.. ..T....... .....C.A.. C.......C. .C........ ........G. A......... A.....
hirsutum12      ..T...G... .A.....G.. .G.....G.. .......... .....C.A.. .......... .C........ ...G....G. .......... ...G..
```

Fig. 1. Continued.

were encountered in two spacer sequences (*raimondii6* and *hirsutum6*).

Two of our 99 clones, *hirsutum6* and *hirsutum10,* are unusually divergent from each other and from the remaining 5S rDNA sequences. Coding regions of these sequences show lower similarity to each other and to other *Gossypium* 5S sequences than exhibited by all other pairwise comparisons (discussed in the following section). In addition, based upon the number of substitutions in the 5S gene, these sequences appear to have experienced accelerated rates of sequence evolution. These features suggest that *hirsutum6* and *hirsutum10* are pseudogenes, although we point out the absence of explicit criteria for determining (from DNA sequence

Table 2. Polymorphism (p_n) and diversity (π) values for *Gossypium* 5S rDNA sequences[a]

		5S gene		Spacer		Entire repeat	
Taxon	*N*	p_n	π	p_n	π	p_n	π
C-genome							
G. robinsonii	4	0.123	0.078	0.073	0.052	0.093	0.062
A-genome/subgenome							
G. arboreum	5	0.141	0.060	0.096	0.057	0.107	0.059
G. herbaceum	5	0.033	0.013	0.051	0.020	0.044	0.018
G. hirsutum	5	0.157	0.069	0.147	0.059	0.151	0.063
G. tomentosum	6	0.157	0.069	0.125	0.044	0.138	0.054
G. barbadense	7	0.149	0.053	0.119	0.037	0.131	0.044
G. mustelinum	5	0.157	0.081	0.096	0.048	0.121	0.061
2(A_2D_1)	5	0.124	0.056	0.096	0.037	0.104	0.041
Overall mean		0.131	0.057	0.109	0.043	0.118	0.049
A-genome diploids		0.087	0.037	0.090	0.038	0.089	0.039
Allopolyploid A-subgenomes		0.155	0.066	0.122	0.045	0.135	0.053
D-genome/subgenome							
G. raimondii	6	0.074	0.027	0.214	0.077	0.158	0.059
G. hirsutum	2	0.076	0.066	0.069	0.065	0.072	0.065
G. hirsutum (Dψ)	2	0.223	0.248	0.148	0.148	0.178	0.188
G. mustelinum	5	0.157	0.072	0.159	0.069	0.158	0.070
2(A_2D_1)	4	0.074	0.040	0.060	0.031	0.066	0.035
G. gossypioides	4	0.132	0.071	0.289	0.154	0.226	0.120
G. laxum	3	0.025	0.022	0.093	0.055	0.066	0.042
G. lobatum	5	0.116	0.064	0.071	0.029	0.089	0.038
G. aridum	2	0.066	0.066	0.094	0.094	0.083	0.083
G. schwendimanii	4	0.083	0.041	0.115	0.066	0.102	0.056
G. klotzschianum	4	0.066	0.039	0.072	0.038	0.070	0.038
G. davidsonii	2	0.058	0.058	0.067	0.067	0.063	0.063
G. thurberi	3	0.099	0.064	0.132	0.091	0.119	0.080
G. trilobum	3	0.050	0.039	0.077	0.053	0.066	0.047
G. turneri	3	0.050	0.033	0.044	0.032	0.046	0.032
G. harknessii	4	0.107	0.059	0.071	0.039	0.086	0.049
G. armourianum	2	0.083	0.083	0.033	0.033	0.052	0.052
Overall mean		0.103	0.064	0.127	0.067	0.117	0.066
D-genome diploids		0.078	0.051	0.094	0.063	0.088	0.058
Allopolyploid D-subgenomes		0.152	0.101	0.126	0.078	0.136	0.090

[a] Nucleotide variability measures are partitioned into values for 5S genes, spacers, and entire 5S rDNA repeats. Polymorphism values are expressed as the proportion of nucleotide positions that are variable. Nucleotide diversity values represent the mean proportion of nucleotide differences among all sequences in a single array from an individual. *N* = the number of clones sequenced per taxon

data) whether rRNAs are functional in translational machinery. Since these two cloned sequences are clearly different from all other sequences, they are considered separately (under the name *G. hirsutum-D*ψ) from D-subgenomic sequences of *G. hirsutum* in subsequent analyses.

Intra-Individual Sequence Variation

The sequence data of Fig. 1 demonstrate that 5S rDNA sequences are highly polymorphic in *Gossypium,* not only between species but also within individual plants. Even though we sampled only a small fraction of the existing repeats from any single genome (two to ten clones per individual), all sequences were unique. Polymorphism appears to be partitioned approximately equally between the 5S gene and the intergenic spacer, although variation is not uniformly distributed across all nucleotides.

A mean of 12% of nucleotide positions are polymorphic within individual arrays of *Gossypium* species, both for genes and spacers, although there is considerable variance around this mean (Table 2). Intra-individual values of p_n range from 0.033 (for *G. herbaceum*) to 0.223 (*G. hirsutum D*ψ) for the 5S gene and from 0.033 (*G. armourianum*) to 0.289 (*G. gossypioides*) for the intergenic spacer (Table 2). Overall, the least polymorphic 5S sequences are from *G. herbaceum* (p_n = 0.044) and *G. turneri* (p_n = 0.046). At the other extreme are the putative pseudogenes from *G. hirsutum* (p_n = 0.223 and 0.148 for the gene and spacer, respectively; p_n = 0.178 overall) and the sequences from *G. gossypioides* (p_n = 0.132 and 0.289 for the gene and spacer, respectively; p_n = 0.226 overall). Despite the wide range of p_n values observed among species and the overall equivalence of polymorphism in the gene and spacer regions, there does not appear to be a correlation between p_n values for the gene and values for the spacer (r^2 = 0.17).

Because p_n tracks the proportion of polymorphic positions without regard to frequency, the measure may be influenced by sampling intensity. To provide estimates of polymorphism that are less biased with respect to sample size, we calculated nucleotide diversity (π, Nei 1987), which is numerically equivalent to the mean number of nucleotide differences per site between all pairs of sequences. Table 2 shows π within species for 5S genes, spacers, and entire repeats. In general, π values parallel the patterns of sequence variation revealed by p_n. Intra-individual values for π range from a low of 0.018 in sequences from *G. herbaceum* to a high of 0.188 for repeats from the putative *G. hirsutum* pseudogenes. Excluding these pseudogene sequences, the highest nucleotide diversity is observed in *G. gossypioides* (0.120), as was the case for p_n. For those taxa where only two sequences were sampled (*G. aridum, G. davidsonii, G. armourianum*), π is identical to p_n. Overall mean nucleotide diversity for 5S genes is significantly higher than for spacer sequences in the A-genome ($\pi = 0.057$ vs 0.043; probability from one-tailed *t*-test = 0.02), but estimates were nearly identical for the two regions from the D-genome repeats ($\pi = 0.064$ vs 0.067 for genes and spacers, respectively; $p = 0.41$). For entire repeats, nucleotide diversity is higher in D-genome than A-genome diploids (0.058 vs 0.039). Within each genome group, no particular pattern of nucleotide diversity is apparent.

We estimated p_n and π separately for sequences from each of the two subgenomes of the allopolyploid species and the synthetic allopolyploid $2(A_2D_1)$ (Table 2). On average, both p_n and π values for 5S genes in natural allopolyploids are slightly higher than values obtained for the spacer sequences of the same repeats, although not significantly so. Approximately the same proportion of nucleotides are polymorphic (13.5%) in repeats from both subgenomes, although nucleotide diversity in the D-like sequences is nearly double (0.090) that of the A-like sequences (0.053). Nucleotide diversities in the A- and D-subgenomic sequences from the synthetic allopolyploid $2(A_2D_1)$ were approximately equivalent (0.041 vs 0.035 overall).

Interspecific Sequence Variation

Excluding the potentially biasing *G. hirsutum-D*ψ sequences, only 17.7% of the 5S rDNA nucleotides are conserved across the 20 taxa; these include 28 of 122 nucleotides (23.0%) in the 5S gene and 28 of 194 (14.4%) aligned nucleotides in the spacer region (Fig. 1). Among the unanticipated observations of this high degree of sequence polymorphism are the cases where mean sequence polymorphism and nucleotide diversity for 5S genes and spacers are higher from *single individuals* than they are in comparisons between species. Nucleotide diversity for spacer sequences from *G. thurberi* ($\pi = 0.091$), for example, is higher than that estimated from comparisons of spacer sequences of *G. thurberi* with those from *G. klotzschianum, G. trilobum, G. turneri, G. armourianum,* and *G. harknessii* (Table 3). In general, this effect appears to be related to the degree of evolutionary divergence, *viz.,* as time since organismal divergence increases, fixed interspecific differences in 5S rDNAs increasingly overwhelm interrepeat polymorphisms within individual arrays. This effect is most pronounced for spacer regions, where mean intergenomic distances are always larger than intra-individual distances (cf. Tables 2 and 3). In contrast, there are a number of cases where intergenomic distances are equivalent to or lower than intra-individual distances for 5S genes (e.g., D-subgenomic sequences from *G. mustelinum* vs nearly all A-type sequences).

5S Copy Number in Gossypium

Copy number varies over twentyfold among the species sampled, from approximately 1,150 copies per 2C genomic equivalent in the D-genome diploid *G. gossypioides* to approximately 23,500 copies in the allopolyploid *G. barbadense* (Table 4). In the A-genome, the two closely related species *G. arboreum* and *G. herbaceum* differ twofold in 5S rDNA copy number, with an average of 5,500 copies. Copy-number variation is even greater in the D-genome species, with a mean of 4,500 copies but a range that varies from 1,150 in *G. gossypioides* to 10,300 in *G. davidsonii.* As with the A-genome species, closely related D-genome species often vary severalfold in the number of 5S genes. For example, copy number varies from approximately 1,700 to 5,800 within the distinctive and closely related (Fryxell 1979, 1992; Wendel and Albert 1992) Mexican arborescent species that comprise the taxonomic subsection *Erioxylum (G. aridum, G. laxum, G. lobatum, G. schwendimanii).* In allopolyploid species, there is a twofold range in copy-number estimates, from 11,200 (in *G. hirsutum*) to 23,500 (in *G. barbadense*). Three of the four allopolyploid species examined (all but *G. hirsutum*) have approximately equal numbers (22,000–23,500) of 5S genes.

We expected that 5S rDNA copy number of the synthetic allopolyploid $2(A_2D_1)$ would be approximately additive with respect to its A-genome and D-genome diploid progenitors. We found that additivity does not hold for the mean copy number, but does when 95% confidence intervals are considered, i.e., summing copy numbers of parents A_2 (5,900–9,000) and D_1 (1,850–2,300) gives an expected range of 6,750–11,300 copies, compared to an observed 95% CI of 10,500–17,500.

Phylogenetic Analysis

We conducted parsimony and distance analyses on 5S rDNA data sets that were treated several ways. In one series of analyses, all 99 sequences were included; this

Table 3. Mean differences among 5S gene (above diagonal) and spacer (below diagonal) sequences for diploid and allotetraploid *Gossypium* species[a]

	1	2	3	4	5	6	7	8	9	10	11	12	13
1. *G. robinsonii*	—	0.077	0.052	0.072	0.076	0.077	0.086	0.071	0.081	0.058	0.071	0.067	0.084
2. *G. arboreum*	0.277	—	0.038	0.063	0.070	0.064	0.076	0.053	0.067	0.045	0.059	0.054	0.072
3. *G. herbaceum*	0.278	0.061	—	0.035	0.045	0.044	0.054	0.035	0.042	0.019	0.035	0.027	0.044
4. *G. barbadense*	0.267	0.057	0.038	—	0.061	0.061	0.071	0.053	0.064	0.042	0.051	0.051	0.068
5. *G. tomentosum*	0.260	0.062	0.044	0.043	—	0.067	0.081	0.065	0.074	0.051	0.061	0.060	0.076
6. *G. hirsutum* (A)	0.273	0.071	0.049	0.054	0.057	—	0.078	0.065	0.068	0.048	0.055	0.056	0.076
7. *G. mustelinum* (A)	0.259	0.067	0.043	0.048	0.053	0.060	—	0.073	0.085	0.060	0.067	0.070	0.088
8. 2(A_2D_1) (A)	0.269	0.049	0.048	0.045	0.052	0.062	0.052	—	0.068	0.038	0.047	0.053	0.084
9. *G. gossypioides*	0.281	0.281	0.272	0.276	0.274	0.281	0.271	0.271	—	0.051	0.062	0.057	0.075
10. *G. raimondii*	0.277	0.298	0.282	0.290	0.282	0.297	0.283	0.285	0.191	—	0.045	0.038	0.051
11. *G. hirsutum* (D)	0.271	0.287	0.268	0.281	0.271	0.284	0.269	0.273	0.191	0.087	—	0.063	0.052
12. *G. mustelinum* (D)	0.264	0.285	0.270	0.279	0.269	0.282	0.266	0.270	0.197	0.091	0.070	—	0.066
13. *G. schwendimanii*	0.250	0.271	0.260	0.266	0.262	0.273	0.261	0.261	0.192	0.130	0.127	0.130	—
14. *G. aridum*	0.257	0.280	0.269	0.275	0.269	0.275	0.270	0.270	0.199	0.144	0.138	0.144	0.080
15. *G. lobatum*	0.252	0.276	0.264	0.271	0.266	0.273	0.267	0.266	0.187	0.133	0.130	0.135	0.080
16. *G. laxum*	0.266	0.291	0.279	0.285	0.281	0.282	0.279	0.279	0.206	0.150	0.148	0.150	0.105
17. *G. klotzschianum*	0.245	0.254	0.242	0.249	0.243	0.252	0.243	0.241	0.167	0.090	0.089	0.093	0.109
18. *G. davidsonii*	0.264	0.273	0.262	0.268	0.264	0.271	0.263	0.261	0.186	0.108	0.100	0.109	0.121
19. *G. trilobum*	0.263	0.268	0.254	0.262	0.255	0.268	0.257	0.255	0.180	0.100	0.097	0.103	0.118
20. *G. thurberi*	0.272	0.269	0.258	0.259	0.255	0.266	0.256	0.256	0.188	0.110	0.107	0.113	0.126
21. 2(A_2D_1) (D)	0.256	0.252	0.239	0.247	0.242	0.252	0.241	0.239	0.170	0.090	0.084	0.090	0.107
22. *G. turneri*	0.259	0.262	0.247	0.254	0.248	0.260	0.247	0.248	0.172	0.080	0.079	0.083	0.089
23. *G. harknessii*	0.257	0.260	0.249	0.256	0.251	0.259	0.249	0.248	0.173	0.084	0.079	0.093	0.092
24. *G. armourianum*	0.256	0.263	0.250	0.258	0.253	0.261	0.276	0.250	0.167	0.074	0.073	0.076	0.094

[a] Each entry represents the mean proportion of nucleotide differences between all sequences involved in the comparison. Two lines and columns are listed for allopolyploid species for which sequences were recovered from both the A- and D-subgenomes

data set was analyzed using sequences from the 5S gene alone (121 aligned characters), the intergenic spacer alone (195 characters), and the entire 5S repeat (316 characters). We explored several alternative gap codings and found that these treatments did not significantly alter the topologies obtained. In a second series of analyses, we performed parsimony and distance-based analyses on data sets of reduced dimensionality, generated by computing "consensus" sequences for each diploid species and for sequences recovered from each subgenome of each allopolyploid species. To accomplish this, positions that were polymorphic within a species (or within a subgenome) were coded using appropriate ambiguity coding. The resulting data set contained 25 rows, consisting of 1 outgroup (*G. robinsonii*), 7 "A-genome" repeat types (2 from the 2 A-genome diploids examined, 4 from the allopolyploids included in the study, and 1 from the synthetic allopolyploid 2[A_2D_1]), and 17 "D-genome" repeat types (13 from D-genome diploids, 3 from tetraploid subgenomes and 1 from 2[A_2D_1]). This 25 by 316 matrix was similarly analyzed using the same gap coding alternatives as described above.

Parsimony analysis of the 5S gene sequences for the original 99 sequences resulted in more than 3,000 equally most-parsimonious trees, at which time the program was stopped and a strict consensus was computed. Each constituent tree had a length of 299 and a relatively low retention index (RI = 0.54). Little resolution was retained in the strict consensus of the shortest trees; nearly all sequences form an unresolved "rake" at the base of the tree (Fig. 2B). The exceptions are two poorly supported clades, one uniting sequences from the phylogenetically distant D-genome species *G. turneri* and *G. schwendimanii* and the other joining sequences from the more closely related species *G. aridum* and *G. lobatum.* Upon decay analysis, however, both of these clades collapse in trees that are only a single step longer than the most parsimonious trees. It is noteworthy that these analyses failed to recover "clades" of sequences from individual species, even within the outgroup species *G. robinsonii,* which is arbitrarily shown as monophyletic in Fig. 2B.

In contrast to a near-absence of phylogenetically useful information in the 5S gene sequences, considerable resolution is evident in parsimony trees from the intergenic spacer (Fig. 2A). As shown in the strict consensus of the 3,000 shortest trees saved (each length 576 and with a RI of 0.87), each genomic group is monophyletic and is strongly supported by decay analysis (Fig. 2A). Within the D-genome clade (Fig. 2A), spacer sequences are resolved into seven well-supported subclades that are generally concordant with previous phylogenetic results (DeJoode 1992; Wendel and Albert 1992; Wendel et al. 1995b). In contrast, A-genome spacer sequences are not resolved into taxonomically meaningful clades, although sequences from individual species often form higher

Table 3. Continued

14	15	16	17	18	19	20	21	22	23	24
0.095	0.071	0.056	0.065	0.075	0.064	0.080	0.066	0.058	0.074	0.085
0.034	0.056	0.040	0.049	0.060	0.046	0.066	0.053	0.045	0.059	0.074
0.024	0.032	0.018	0.027	0.037	0.024	0.040	0.031	0.020	0.039	0.051
0.037	0.055	0.037	0.048	0.058	0.047	0.063	0.048	0.043	0.060	0.068
0.060	0.065	0.049	0.057	0.064	0.057	0.070	0.058	0.051	0.071	0.071
0.024	0.061	0.044	0.053	0.058	0.053	0.064	0.055	0.047	0.066	0.077
0.100	0.068	0.052	0.064	0.070	0.064	0.079	0.065	0.060	0.072	0.065
0.034	0.054	0.036	0.046	0.059	0.046	0.070	0.047	0.043	0.054	0.071
0.072	0.061	0.048	0.057	0.062	0.052	0.067	0.061	0.048	0.069	0.084
0.032	0.039	0.024	0.035	0.044	0.032	0.051	0.039	0.028	0.044	0.056
0.063	0.052	0.037	0.046	0.037	0.043	0.064	0.044	0.039	0.057	0.067
0.082	0.068	0.050	0.060	0.065	0.058	0.072	0.055	0.053	0.068	0.076
0.053	0.045	0.033	0.043	0.054	0.040	0.053	0.050	0.035	0.055	0.068
—	0.059	0.054	0.062	0.071	0.057	0.071	0.068	0.054	0.072	0.079
0.056	—	0.034	0.046	0.056	0.042	0.058	0.050	0.040	0.054	0.072
0.114	0.100	—	0.028	0.042	0.029	0.043	0.030	0.024	0.040	0.055
0.126	0.114	0.127	—	0.046	0.039	0.054	0.041	0.036	0.049	0.063
0.138	0.130	0.144	0.054	—	0.047	0.064	0.048	0.044	0.062	0.071
0.135	0.118	0.132	0.072	0.090	—	0.050	0.040	0.032	0.048	0.064
0.141	0.131	0.133	0.083	0.100	0.071	—	0.055	0.049	0.067	0.083
0.125	0.112	0.120	0.063	0.082	0.044	0.059	—	0.037	0.052	0.062
0.112	0.101	0.112	0.054	0.077	0.066	0.079	0.055	—	0.046	0.058
0.115	0.104	0.121	0.059	0.080	0.071	0.082	0.061	0.037	—	0.074
0.112	0.102	0.119	0.052	0.071	0.065	0.079	0.055	0.033	0.037	—

Table 4. Mean 5S rDNA copy-number estimates in *Gossypium*, based on slot-blot hybridization and phosphorimaging[a]

Taxon	*N*	Copy number	95% CI	CV
C-genome				
G. robinsonii	4	2,940	2,121–3,755	17.5%
A-genome				
G. herbaceum	8	3,415	2,609–4,222	28.2%
G. arboreum	9	7,550	5,900–9,037	21.0%
D-genome				
G. gossypioides	8	1,145	812–1,477	34.8%
G. trilobum	8	1,420	1,101–1,737	26.8%
G. aridum	8	1,685	492–2,880	84.7%
G. thurberi	8	2,070	1,862–2,280	12.1%
G. laxum	8	2,760	2,289–3,230	20.4%
G. armourianum	8	4,205	3,301–5,105	25.7%
G. harknessii	8	4,385	3,015–5,750	37.3%
G. schwendimanii	8	4,615	3,964–5,266	16.9%
G. raimondii	8	4,730	3,838–5,623	22.6%
G. lobatum	8	5,830	4,048–7,614	36.6%
G. klotzschianum	8	6,930	5,509–8,346	24.5%
G. turneri	7	8,495	6,934–10,054	19.9%
G. davidsonii	8	10,280	8,518–12,039	20.5%
AD-genome				
G. hirsutum	8	11,190	8,154–14,228	32.5%
2(A_2D_1)	8	14,015	10,503–17,522	30.0%
G. mustelinum	8	21,845	15,007–28,684	37.4%
G. tomentosum	8	22,290	18,822–25,765	18.6%
G. barbadense	8	23,515	20,100–26,929	17.4%

[a] *N* = the number of replicate experiments; 95% CI = 95% confidence interval; CV = coefficient of variation

level organization (e.g., clades of *tomentosum1, tomentosum6,* and *tomentosum10,* and of *herbaceum1, herbaceum2,* and *herbaceum5*).

Analysis of 3,000 minimal length trees derived from entire 5S rDNA repeats (gene + spacer) resulted in a strict consensus topology (not shown) that is identical to that illustrated for the intergenic spacer alone (Fig. 2A). Due to the additional and ''noisy'' characters from the 5S gene sequences, these trees are longer (978 steps) and show slightly more homoplasy (RI = 0.80) than trees generated from intergenic spacer sequences alone.

Our final parsimony analyses utilized ''consensus'' sequences for each diploid species and sets of sequences from individual subgenomes of the allopolyploids. The topologies of strict consensus trees generated from spacer sequences alone (3,000 trees saved, each of length 121 and a RI of 0.95) and from entire 5S repeats (5S gene + spacer; 3,000 trees saved, each of length 159 and a RI of 0.94) were identical to each other and had identical decay values for most clades. The consensus tree from the spacer sequence data set is shown in Fig. 3, which shows that even when the data are reduced to consensus sequences, genomic identity is retained, with the formation of two strongly supported clades corresponding to the A and D genomes. In general, trees generated from consensus sequences exhibited slightly less resolution than cladograms derived from individual sequences.

Distance-based methods of phylogenetic analysis produced trees that were concordant with those derived from parsimony analysis in that all clades observed in strict

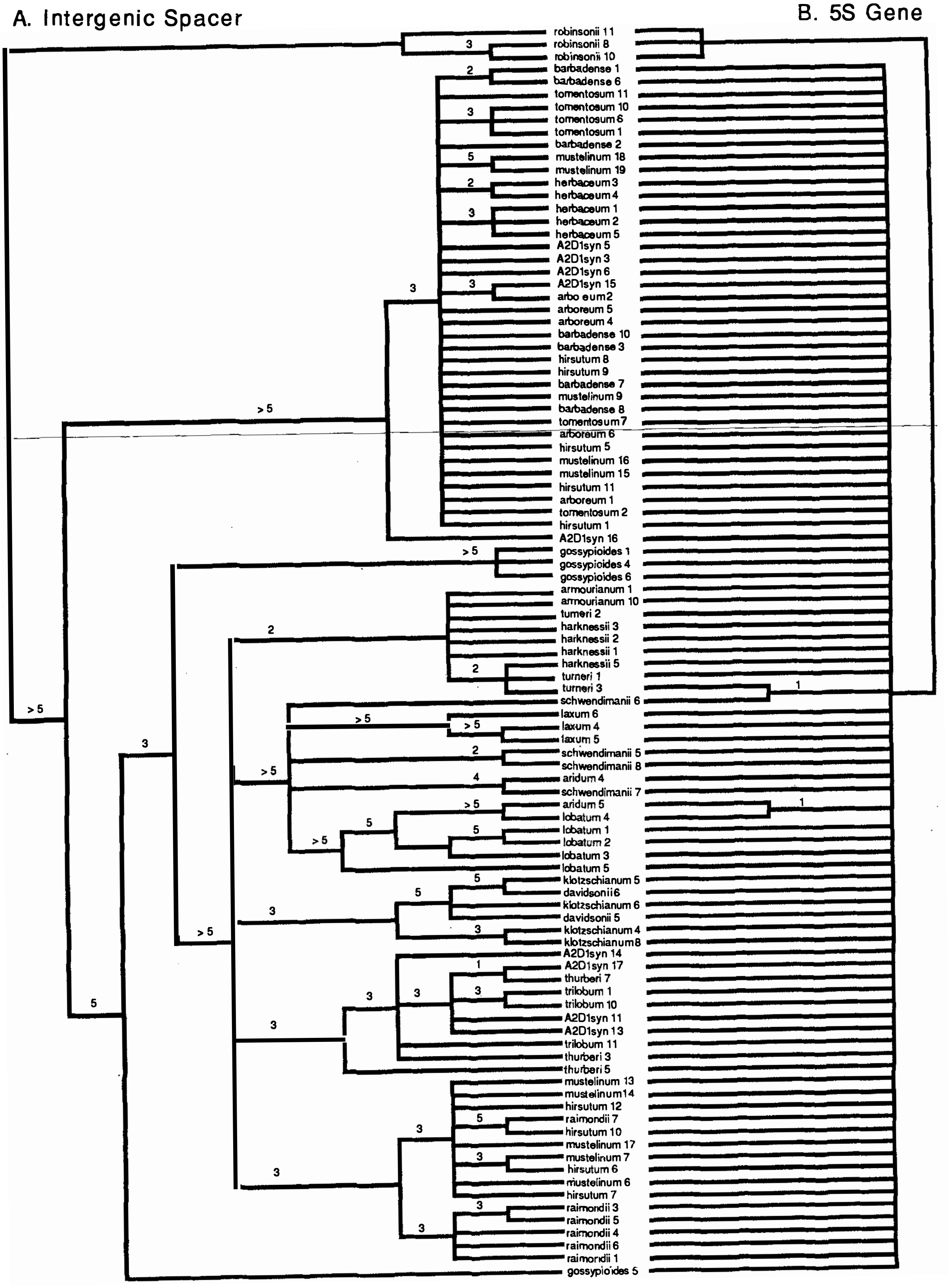

Fig. 2. Consensus gene trees resulting from maximum parsimony analysis of 99 cloned 5S rDNA sequences from *Gossypium.* Topologies shown are the strict consensus of 3,000 trees recovered from heuristic searches. *Numbers* above branch segments indicate the number of additional steps that are required for each resolved clade to collapse ("decay values"). For example, in trees two steps longer than the most parsimonious, the clade consisting of *schwendimanii5* and *schwendimanii8* is no longer supported. **A** Consensus tree based on sequence data from the intergenic spacers alone (length of each constituent tree = 576 steps; consistency index = 0.55; retention index = 0.87). **B** Consensus tree based on sequence data from the 5S genes alone (tree length = 299 steps; CI = 0.60; RI = 0.54). Both trees are rooted with sequences from the outgroup taxon *G. robinsonii,* which resolved as monophyletic in the tree based on 5S spacer sequences (Fig. 2A) but was constrained to be monophyletic in the tree based on 5S gene sequences (Fig. 2B).

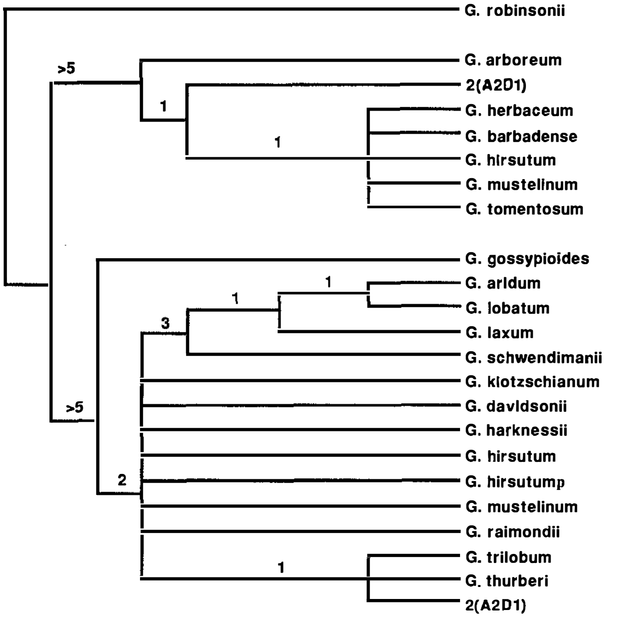

Fig. 3. Consensus tree resulting from maximum parsimony analysis of 23 5S rDNA spacer sequences from *Gossypium.* For each diploid species, a consensus sequence was computed prior to phylogenetic analysis, as described in the text. Similarly, consensus sequences were computed separately for sequences from each subgenome of the allopolyploids. The topology shown, rooted with the outgroup taxon *G. robinsonii,* is the strict consensus of 3,000 minimum length trees recovered from heuristic searches. *Numbers* above branch segments indicate the number of additional steps that are required for each clade to collapse. Each constituent tree had a length of 121 steps, a consistency index of 0.87, and a retention index of 0.95.

consensus trees generated by parsimony methods are supported by neighbor-joining trees (Fig. 4). In general, analysis of spacer sequences alone (as opposed to the entire 5S repeats) provides the greatest degree of resolution, in the sense that more sequences form species-specific clades. It also appears that the high level of homoplasious polymorphism contained within 5S gene sequences acts to reduce intersequence distances, which shortens internode lengths, thereby reducing confidence in the resolution obtained. In the tree shown (Fig. 4), groups within the D-genome clade appear to be more clearly resolved (with longer internode lengths) than sequences within the A-genome, as was the case with parsimony analysis. Among the more intriguing aspects of the neighbor-joining results are relationships revealed between individuals of closely related species pairs, such as *G. thurberi/G. trilobum* and *G. davidsonii/G. klotzschianum.* Both of these species pairs form well-defined clusters that are distinct from other taxa, yet within each cluster species-specific groups are not formed. Included in the alternative explanations for this pattern are high mutability of particular nucleotide positions (homoplasies that mimic synapomorphies at this level) and retention of 5S polymorphisms that are older than the speciation event that separated the species.

Discussion

Gossypium 5S rDNA Structure and Organization

The 5S rDNA arrays in all *Gossypium* species examined exhibit a conventional organization of tandemly repeated 5S genes and intergenic spacers. Fluorescent *in situ* hybridization work demonstrates that these arrays occupy a single centromeric location in A-genome and D-genome diploid species (R. Hanson and D. Stelly, pers. comm.) and two corresponding loci in the AD-genome allopolyploids (Crane et al. 1993). *Gossypium* 5S genes and spacers range from 121–122 and 175–191 nucleotides in length, respectively. Totaling 296 to 311 bp, these repeats are among the shortest known in plants, with only three reports of smaller repeat units (*Datisca glomerata* = 219 bp; *Gleditsia triacanthos* = 278 bp; *Gymnocladus dioicus* = 215 bp; Gottlob-McHugh et al. 1990). Most variation in the length of the *Gossypium* 5S rDNA repeats is attributable to the 7-bp (distinguishing A-genome and D-genome from C-genome taxa) and 8-bp (A-genome) genome-specific indels, although minor intra-individual length variation is also evident; the latter appears to result from contraction/expansion of the T-rich region downstream of the 5S gene coding region. Three highly conserved hallmarks characterize *Gossypium* 5S genes and spacers: (1) the *Bam*HI site used for cloning (nucleotides 30–35, Fig. 1), which is a conserved feature of land plants (Sastri 1992); (2) a pentanucleotide "TATRA" motif 22–26 bp upstream of the 5S gene (nucleotides 291–295), which is commonly observed in plants (*Acacia,* Playford et al. 1992; *Arabidopsis,* Campell et al. 1992; *Bromus,* Sastri et al. 1992; *Glycine,* Gottlob-McHugh et al. 1990; *Lupinis,* Rafalski et al. 1982; *Sinapis,* Capesius 1991; *Vigna,* Hemleben and Werts 1988; *Zea,* Sastri et al. 1992; the *Triticeae,* Cox et al. 1992, Kellogg and Appels 1995) and has been implicated in transcription initiation of 5S (Tyler 1987; Sharp and Garcia 1988) and other class III genes (White 1994); and (3) a "TTTTATAT" motif immediately downstream of the gene at nucleotides 126–133, which is thought to facilitate efficient transcription termination (Korn 1982).

5S Copy Number Is Evolutionarily Labile

The number of 5S rDNA repeats per genome varies over twentyfold among species (Table 4), thereby demonstrating that arrays have expanded and contracted since the origin of the genus. Specific examples of array expansion or contraction are evident, despite the sizable errors associated with most estimates. These inferences are dependent on correct diagnoses of ancestral conditions, which are in turn dependent on the organismal phylogeny (DeJoode 1992; Wendel and Albert 1992; Wendel et

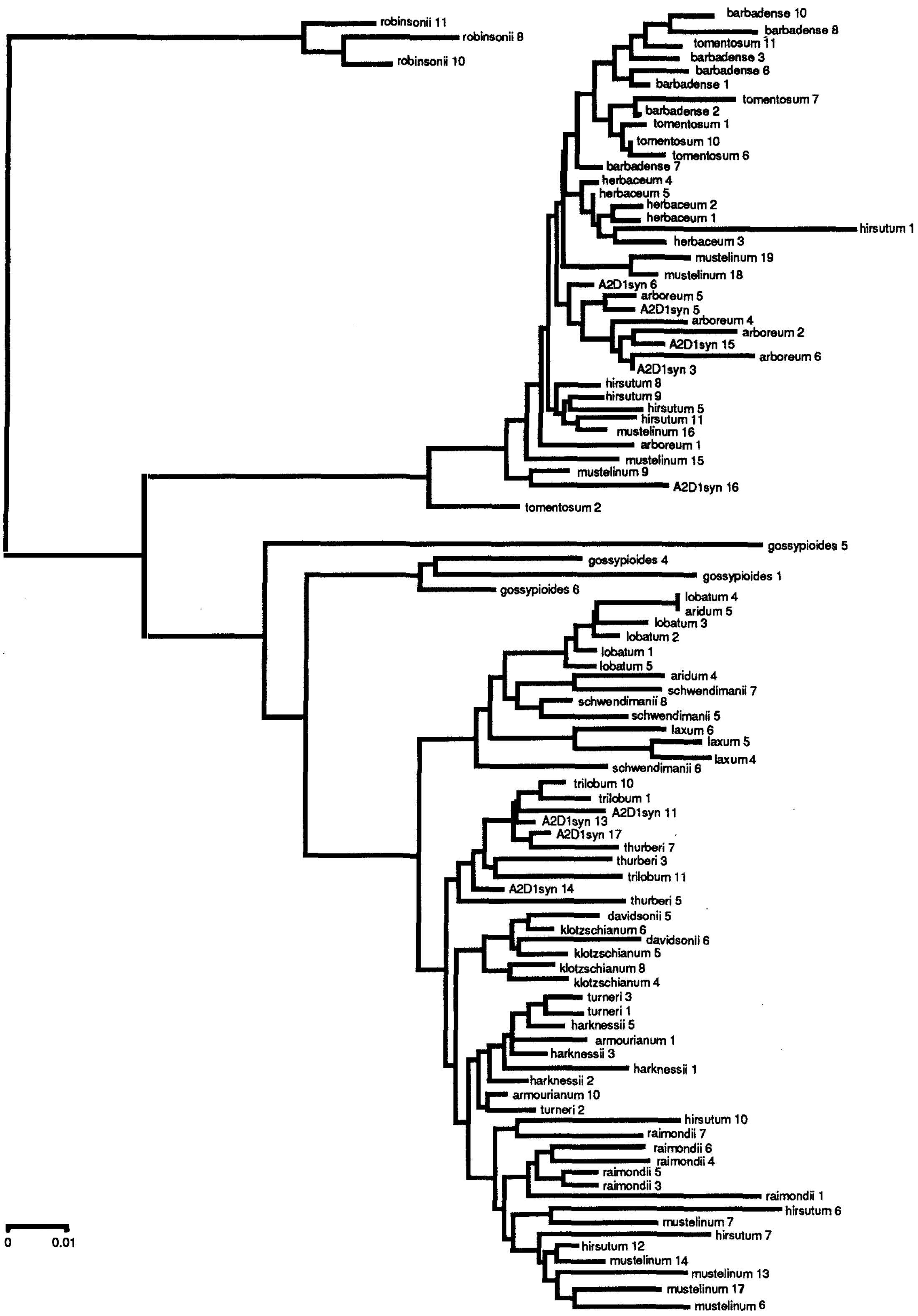

Fig. 4. Neighbor-joining tree based on Kimura two-parameter distances between 99 cloned 5S rDNA spacer sequences from *Gossypium*. The tree is rooted with sequences from the outgroup taxon *G. robinsonii*. Branch lengths are drawn to scale (shown at bottom).

al. 1994, 1995b). For example, if the common ancestor of the D-genome species had a similar number of 5S rDNA copies to that which both the A-genome diploids (5,500 copies/2C) and the C-genome outgroup species *G. robinsonii* (3,000 copies/2C genome) had, then: (1) the species pair *G. trilobum* and *G. thurberi* experienced a lineage-specific decrease in copy number to an average of 1,700 copies/2C genome and (2) the species pair of *G. davidsonii* and *G. klotzschianum* experienced a lineage-specific increase to approximately 8,600 copies per 2C genome equivalent. Similarly, in the monophyletic subsection *Erioxylum,* mean copy number is 3,700 copies/2C genome, which is lower than for two members of the subsection (*G. schwendimanii* = 4,600; *G. lobatum* = 5,800), but higher than for the other two included species (*G. laxum* = 2,800; *G. aridum* = 1,700). While considerable variation exists in our copy-number estimates, comparisons of 95% confidence intervals (Table 4) and the results of *t*-tests (not shown) show that many of these differences are statistically significant. This shows that both array expansion and contraction can occur within a relatively brief evolutionary time frame.

In contrast to the additivity observed for the synthetic allopolyploid, 5S copy numbers from the putative progenitors of natural allopolyploids (*G. herbaceum* = 3,400, *G. raimondii* = 4,750; $\sum$ = 8,150) add up to less than half of the average copy number for the AD-genome species. These numbers range from 11,200/2C genome in *G. hirsutum* to over 22,000 copies in the other polyploids. Even when we consider experimental error and interspecific variation (Table 4), 5S rDNA copy number clearly is not additive in the allopolyploids. This suggests, but does not prove, that 5S rDNA arrays have expanded in the allopolyploids since their formation.

A final comment with respect to 5S copy number is stimulated by the observation of relatively low copy numbers in *G. gossypioides* and *G. aridum.* Of the diploid species included in this study, only these two are known to have evolutionary histories that include episodes of interspecific hybridization and introgression (DeJoode 1992; Wendel and Albert 1992; Wendel et al. 1995b). In the case of *G. aridum,* the cytoplasmic parent was similar to present-day *G. klotzchianum* (with 6,950 5S rDNA copies/2C genome) and the paternal parent was similar to members of the present-day subsection *Erioxylum* (with a mean of 3,700 copies/2C genome). Although copy-number estimates for *G. aridum* have a large error (Table 4), the mean of 1,700 copies/2C genome is much lower than that of either parental lineage. Similarly, the values obtained for the intergenomic derivative *G. gossypioides* (1,150 copies/2C genome) are substantially lower than for all other species examined. In this respect it is noteworthy that Zimmer et al. (1988, p. 1134) reported that in hybrids between *Zea mays* and *Z. diploperennis,* "teosinte-specific genes (rDNA and 5S DNA) are underrepresented in the F_1 hybrids analyzed." Given our small sample of two species, the association between reticulation and 5S rDNA copy-number reduction may be coincidental. However, selection may operate to reduce copy number in hybrids as a means of rapidly eliminating excess sequence variation within an array, perhaps as a necessary component of optimizing 5S rDNA expression and ribosomal composition or function. If this hypothesis is true, our observation of 5S rDNA copy-number reduction may be a significant aspect of the stabilization of hybrid evolutionary products. Clearly, other natural and synthetic interspecific hybrids and later-generation segregates need to be screened for 5S (and perhaps other repetitive DNA) copy number to evaluate whether contraction of repeated sequence arrays is a common consequence of interspecific hybridization.

5S Sequence Evolution and Phylogeny Reconstruction

Several authors have examined the value of 5S rRNA genes (Wheeler and Honeycutt 1988; Steele et al. 1991; Halanych 1991; Vawter and Brown 1993) and spacer sequences (Scoles et al. 1988; Baum and Appels 1992; Kellogg and Appels 1995) for phylogeny reconstruction. Although the 5S gene may provide useful information, it has been applied mostly to relatively ancient divergences, such as between major clades of prokaryotes (Woese 1987) and between the major eukaryotic phyla (Wheeler and Honeycutt 1988; Steele et al. 1991). In addition, its length (*ca.* 120 bp) places practical limitations on its ability to record evolutionary history. The spacer region, although also reasonably short (100–700 bp in plants; Sastri et al. 1992), has a much higher rate of sequence substitution, and hence it is more likely to provide phylogenetically useful information at lower taxonomic ranks. As with many spacer sequences, however, alignment difficulties are likely to arise as more divergent taxa are included in an analysis, due to the characteristic occurrence of simple repeats (Kanazin et al. 1993) and indels (Cox et al. 1992; this paper).

We evaluated the phylogenetic utility of each region in the *Gossypium* 5S rDNA repeats by using both character-based and distance-based approaches to phylogeny estimation. Analysis of the gene sequences alone resulted in a large number of minimal length trees (>3,000) with low retention indices (RI = 0.54) and a strict consensus with virtually no resolution (Fig. 2B). A notable feature of this tree is that in addition to the absence of cladistic structure among species, there is a complete lack of resolution of sequences from *individual* species. This demonstrates that the 5S gene in *Gossypium,* despite exhibiting reasonably high intraspecific polymorphism (mean p_n = 0.12; π = 0.06; Table 2) and interspecific divergence (28 of 121 nucleotides conserved across taxa; also Table 3), is not phylogenetically useful within *Gossypium.* The lack of phylogenetic information is not due to an absence of variation, which might have been the a priori expectation given the presumed slow rate of sequence evolution in 5S genes. Instead, there is an abundance of sequence variation, but it is highly homoplasious.

A different pattern emerged when the intergenic spacer sequences were analyzed. Although numerous (>3,000) minimal length trees were still found, they each had a high retention index (0.87). In addition, considerable resolution is retained in the strict consensus tree, and clades consisting of sequences from closely related species are often recovered. Also, as shown in Fig. 2A, each diploid genomic group is monophyletic (excluding allopolyploids) and is strongly supported by decay analysis.

There is considerable congruence between the 5S intergenic spacer ''gene tree'' (Fig. 2A) and previous phylogenetic results (DeJoode 1992; Wendel and Albert 1992; Wendel et al. 1995b). Within the D-genome clade, seven well-supported subclades were recovered. The two most basal of these are comprised solely of sequences from the Mexican species *G. gossypioides,* the evolutionary history of which has recently been reviewed (Wendel et al. 1995b). In brief, all sources of evidence, including comparative analysis of chloroplast DNA restriction site data and interspecific fertility relationships, indicate that the sister species of *G. gossypioides* is *G. raimondii.* The sole previous exception to this unanimity consisted of DNA sequence data from the internal transcribed spacer (ITS) region of the 18S-5.8S–26S array, which allied *G. gossypioides* most closely to the A-genome cottons, albeit in a phylogenetically basal position. Wendel et al. (1995b) argued that the incongruence between all other sources of data and the ITS data reveals an ancient hybridization and introgression event between the antecedent of modern *G. gossypioides* and an A-genome ITS. The present results for the 5S rDNA are similar, in that *G. gossypioides* again occupies a phylogenetically basal position, albeit in the D-genome clade rather than the A-genome clade. As hybrid taxa are *expected* to occupy phylogenetically basal positions in cladistic analyses (McDade 1990, 1992), we interpret these results as additional support for the hybridization and introgression hypothesis advanced by Wendel et al. (1995b).

Resolution of the remaining D-genome sequences is into a polytomy that unites five major clades: (1) the arborescent species comprising the Mexican subsection *Erioxylum (G. laxum, G. schwendimanii, G. lobatum, G. aridum);* (2) the Baja/Galapagos Islands species pair (Wendel and Percival 1990) that comprises subsection *Integrifolia (G. klotzschianum, G. davidsonii);* (3) the two members of subsection *Houzingenia (G. thurberi, G. trilobum);* (4) the members of subsection *Caducibracteolata* from Baja, California *(G. turneri, G. harknessii, G. armourianum);* and (5) the clade uniting the D-type repeats from the allopolyploids with all sequences from *G. raimondii.* This last clade provides additional evidence in support of the traditional hypothesis that *G. raimondii* is the best living model of the original D-genome donor to the allopolyploids (Endrizzi et al. 1985; Wendel 1989; but see Wendel et al. 1995b).

Duplicated 5S Arrays Evolve Independently in Allopolyploids

Among the more important results is that different sequences from single allopolyploid species often occur in both the A-genome and D-genome clades (Figs. 1–4). Using sample sizes of nine to ten clones per taxon, we were able to isolate two distinct classes of 5S rDNA sequences from the natural allopolyploids *G. hirsutum* and *G. mustelinum* and from the synthetic allopolyploid $2(A_2D_1)$. These two classes of sequences evidently originated from the two different allopolyploid subgenomes (A and D), as each class shares a high degree of sequence similarity to 5S repeats from the putative subgenome donors. Moreover, in each species both repeat types were detected in nearly equal proportions. This demonstrates that for these species, orthology-paralogy relationships have been retained since allopolyploid formation (1–2 MYBP; Wendel 1989; Wendel and Albert 1992).

The significance of this observation is that it constitutes compelling evidence that *intralocus* concerted evolution has predominated over *interlocus* interactions. This conclusion seems firm, although it is not without precedent; in fact, we are aware of no case where interlocus concerted evolution of 5S rDNA arrays has been demonstrated in plants (Cox et al. 1992; Sastri et al. 1992; Kellogg and Appels 1995). Available information, therefore, suggests that the predominant homogenizing forces acting on 5S ribosomal genes and spacers operate at the level of the individual array.

From previous analyses of the 45S arrays in the same allopolyploid *Gossypium* species (Wendel et al. 1995a), we know that interlocus evolution has homogenized, to near-identity, sequences located on homoeologous chromosomes. The differences in the evolutionary behavior of 5S and 45S arrays indicate that relative to 45S arrays, interlocus interactions among 5S arrays are prohibited or are too infrequent to be detected (cf. Dover 1994; Schlötterer and Tautz 1994). While 5S rRNA and 45S rRNA genes exist as highly repetitive, tandemly arranged arrays, differences exist in both the number and organization of those arrays. Specifically, *Gossypium* allopolyploids have inherited one 5S locus but two major 45S loci from each parent (Crane et al. 1993). Moreover, 5S rDNA loci are located near the centromere whereas 45S rDNA loci occupy telomeric or subtelomeric locations. Under the assumptions that (1) unequal crossing-over is the operative mechanism of interlocus homogenization in *Gossypium,* and (2) unequal crossing-over in centromerically located arrays would lead to unbalanced and presumably inviable gametes, Wendel et al. (1995a) suggested that long-term maintenance of interlocus polymorphism following allopolyploidization is more likely for sequences that are centromeric rather than telomeric in distribution. The 5S results presented here appear to meet that prediction.

Although both A- and D-subgenomic homoeologues were recovered from *G. hirsutum, G. mustelinum,* and the synthetic $2(A_2D_1)$, only the A-subgenome repeat was detected from *G. barbadense* ($N = 7$) and *G. tomentosum* ($N = 6$). There are at least three possible explanations for this observation: (1) that the 5S rDNA locus has been lost from the D-subgenome of these two species; (2) that D-type 5S arrays exist but were missed due to sampling or experimental bias; and (3) that interlocus concerted evolution has converted D-type 5S arrays to A-type only in *G. barbadense* and *G. tomentosum.*

The first alternative, reduction or loss of a 5S array, has been documented for hexaploid wheat (Dvorák 1990) but is contraindicated for *Gossypium* by fluorescent *in situ* hybridization results that reveal two arrays—one on each of the homoeologous chromosomes—for *G. barbadense, G. hirsutum,* and *G. mustelinum* (R. Hanson and D. Stelly, pers. comm.). By extrapolation, these results suggest that loss of the 5S locus is also unlikely in *G. tomentosum.* To discriminate between the alternatives (2) and (3)—sampling vs interlocus concerted evolution—we used the sequence data (Fig. 1) to develop a 24-bp PCR primer (''gapR'' = 5′-TCA-AAT-TAT-TTA-TTT-CAC-AAA-ACG) that hybridizes specifically to D-subgenomic sequences in the region of indel 2 (nucleotides 204–227). This primer, when paired with 5SF, was expected to amplify a 159-bp fragment from D-genome but not from A-genome repeats. Using M13 clones of known genomic origin as templates, this expectation was met, since only clones from D-genome diploids or the D-subgenome of allotetraploids showed the expected PCR fragment (data not shown). When genomic DNAs were used as templates, D-genome diploids and *all* allotetraploids showed the expected amplification product, whereas A-genome diploids yielded no PCR product. These results constitute strong evidence that D-subgenomic repeats (and by inference, D-subgenomic arrays) are present in the genomes of *G. barbadense* and *G. tomentosum,* and that these sequences were not detected in our M13 clones due to sampling or experimental bias. We conclude, therefore, that there is no evidence of interlocus concerted evolution of 5S arrays in *Gossypium* allopolyploids.

Evolution of 5S rDNA: A Balance of Mutational, Homogenizing, and Selective Forces

In *Gossypium,* mean intraindividual nucleotide diversity for 5S genes is nearly identical to the mean diversity found in spacer sequences (0.061 vs 0.060; Table 2). This level of intra-individual polymorphism is approximately equal to that observed in other plant species. A survey of 28 diploid species from the wheat tribe *Triticeae* revealed nucleotide diversity values of 0.00–0.06 for the 5S gene and 0.00–0.11 for spacer sequences (Kellogg and Appels 1995), while lower values were obtained for nine 5S rDNA sequences from *Glycine max* (π = 0.01 for both gene and spacer; Gottlob-McHugh 1990). Despite these data demonstrating intra-individual 5S rDNA polymorphism, concerted evolutionary processes are clearly homogenizing 5S rDNA sequences. In *Gossypium,* this is evidenced by the presence of highly conserved repeat length in diploid and polyploid genomic groups (A = A' = 295–298 bp; D = D' = 301-304 bp; C = 310–314 bp) and the phylogenetic conclusion (Figs. 2–4) that different sequences from single diploid species are, for the most part, more similar to each other than they are to sequences from other species. Neither of these observations is consistent with a mode of evolution in which each individual repeat evolves independently. Similar results have been obtained from phylogenetic analysis of long and short 5S repeats from members of the *Triticeae* (Sastri et al. 1992; Kellogg and Appels 1995). Taken as a whole, these results underscore the apparent contradiction that although 5S genes are subjected to concerted evolutionary forces, they display considerable intra-individual polymorphism. The heterogeneity observed, therefore, must reflect the net effect of opposing and complementary forces that operate on 5S arrays; these include mutation, homogenization, and selection.

The rate at which repeated sequences interact is an important factor in determining the degree of polymorphism maintained in an array. Although rates of concerted evolution can be predicted by models (Nagylaki and Petes 1982; Ohta 1983, 1984, 1990; Ohta and Dover 1983; Nagylaki 1984a,b, 1990; Basten and Ohta 1992), the application of these models to empirical observations can be problematic. At the simplest level, a survey of the amount of polymorphism that is retained across speciation events allows inferences to be made regarding the frequency with which polymorphism is removed from arrays. If concerted evolutionary processes homogenize 5S repeats at rates greater than the rate of speciation, novel mutations are expected to become fixed or removed and sequence polymorphism is expected to be low within species, with the absolute level determined by the severity of the homogenizing forces. Alternatively, if concerted evolutionary events homogenize 5S rDNA at rates equivalent to or slower than the rate of speciation, one expects greater levels of polymorphism within arrays. In addition, since polymorphism can survive through one or more speciation events, a corollary expectation is that closely related species will share 5S rDNA polymorphisms. In our data (Fig. 1), polymorphisms that are shared between closely related species are evident in both the 5S gene and spacer (e..g, nucleotide positions 13, 182, 253). The most parsimonious interpretation of these shared polymorphisms is that they reflect a single mutation in the common ancestor that survived through a speciation event and has escaped homogenization in both daughter species. That these shared polymorphisms are restricted to closely related species leads to the qualitative generalization that fixation rates are approximately equal to rates of speciation in *Gossypium.*

Although this interpretation accounts for the majority of the shared polymorphism in the sequence data, there

Table 5. Tests for equivalent patterns of sequence evolution in 5S genes and spacers[a]

	Number and type of nucleotide differences for each region				
	Fixed		Polymorphic		
Pairwise comparison	5S Gene	Spacer	5S Gene	Spacer	*P*
Within genomes/subgenomes:					
G. herbaceum vs *G. arboreum*	0	3	19	27	0.273
G. herbaceum vs *G. mustelinum* (A)	0	1	20	23	1.000
G. raimondii vs *G. lobatum*	0	13	21	41	0.015
G. raimondii vs *G. mustelinum* (D)	0	2	23	50	0.569
Between genomes:					
G. herbaceum vs *G. lobatum*	0	42	14	18	0.001
G. herbaceum vs *G. raimondii*	0	36	12	39	0.003
G. arboreum vs *G. raimondii*	0	37	22	44	0.001
G. mustelinum (A) vs *G. raimondii*	0	35	24	43	0.001
G. mustelinum (A) vs *mustelinum* (D)	0	32	32	33	0.001

[a] Two-by-two contingency tables were constructed for the taxa shown, where the observed numbers of fixed and polymorphic differences (columns) were tabulated for 5S gene and spacer sequences (rows)

are notable exceptions at positions 49 and 68 of the 5S gene (Fig. 1). At these two positions, polymorphisms occur throughout the genus, reflecting, in both cases, transitional substitutions (purines at position 49, pyrimidines at 68). Position 49 of the 5S gene, for example, is polymorphic in nearly every species examined, with guanine and adenine represented in approximately equal ratios in all species where polymorphism was detected. Considering the amount of time since divergence of C-genome cottons (20–30 million years; Wendel and Albert 1992) from the remainder of the genus, as well as since separation of A- and D-genome lineages from each other (5–10 million years), it is unlikely that this polymorphism reflects the retention of a single ancestral polymorphism that has yet to become fixed or lost. A more likely scenario is that these sites are evolutionary labile, allowing transitional mutations to occur repeatedly during radiation of the genus. An explanation for the absence of transversions may be that these nucleotides occur within the 5S gene; transversions at these two positions, once propagated and homogenized across the 5S array, may alter ribosome function and reduce relative fitness.

This example may be revealing with respect to the forces that govern 5S sequence evolution. Trees based on 5S genes yield unresolved ''rakes,'' whereas those based on spacer sequences contain considerable resolution (Fig. 2). The absence of resolution in trees based on the coding sequences clearly does not reflect a paucity of variation, as diversity and polymorphism are approximately equivalent for genes and spacers. It also seems unlikely that concerted evolutionary processes discriminate between 5S gene and spacer sequences, although no empirical evidence directly eliminates this as a formal possibility. Instead, the impressive difference in resolution between the trees based on 5S genes and those based on spacer sequences reflects a fundamental contrast in the nature of 5S gene and spacer sequence evolution. Specifically, although nucleotide substitutions appear to accumulate in roughly equal proportions in genes and spacers, *fixation* of these differences among repeats in an array is limited to spacer regions.

This conclusion, which emerged from inspection of the sequence data (Fig. 1) and from the list of character-state changes in parsimony-based trees (Fig. 2), was evaluated statistically (after Kellogg and Appels 1995). To do this, we made pairwise comparisons among a subset of taxa for which five or more sequences were generated; in each comparison, we compiled 2 × 2 contingency tables of fixed vs polymorphic differences (columns) in the 5S gene and spacer (rows), and tested for independence of row and column categories by using the two-tailed Fisher exact test (Sokal and Rohlf 1981). The results show that for all comparisons of taxa from separate, well-supported clades, probabilities of independence are less than 0.05 (Table 5). In all cases, it is a deficiency of fixed differences in the 5S gene and a surplus of fixed differences in the spacer that are responsible for the statistical significance. This is most evident in comparisons involving species from the A- vs D-genomes: despite an accumulation of from 32 to 42 fixed differences in the spacer region, not a single fixed difference has evolved in the 5S gene. A consequence of this phenomenon is that there is no phylogenetic content in the 5S genic data: the ample polymorphism that exists resolves only as autapomorphy and homoplasy (Fig. 2).

To extend the analysis beyond these select taxa we computed consensus sequences for all species and for the genus as a whole. This confirmed that not a single mutation has become fixed in the 5S gene during the 20-30-million-year history of the genus (Wendel and Albert 1992). Moreover, the consensus for *Gossypium* is identical to that computed for 152 sequences from the *Triticeae* (Kellogg and Appels 1995), implying that selection

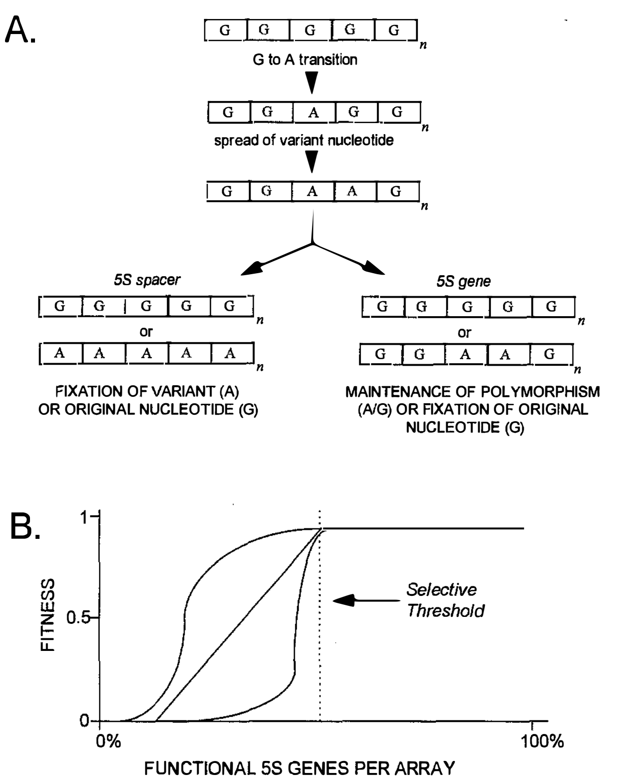
A.
G G G G G n
G to A transition
G G A G G n
spread of variant nucleotide
G G A A G n
5S spacer
G G G G G n
or
A A A A A n
FIXATION OF VARIANT (A) OR ORIGINAL NUCLEOTIDE (G)
5S gene
G G G G G n
or
G G A A G n
MAINTENANCE OF POLYMORPHISM (A/G) OR FIXATION OF ORIGINAL NUCLEOTIDE (G)
B.
1
0.5
0
FITNESS
Selective Threshold
0%
100%
FUNCTIONAL 5S GENES PER ARRAY

Fig. 5. Model illustrating the major features of 5S rDNA evolution in *Gossypium.* **A** Nucleotide substitutions in individual repeats of an array generate polymorphisms. Concerted evolutionary forces may eliminate variants or spread them throughout individual arrays (intralocus), but not between different arrays (interlocus). Most nucleotide positions in the spacer region are presumed to be free to vary, i.e., variants are selectively neutral or near-neutral. Consequently, variant nucleotides can become either fixed or lost. In contrast, most mutations in the 5S gene are presumed to be selectively neutral or near-neutral only when they occur in a subcritical proportion of repeats in an array. Once this threshold frequency is reached, the variant nucleotide becomes disadvantageous due to its effects on 5S transcription or 5S RNA function; thus, the *array* acquires reduced fitness and is selectively removed from the population. As a consequence, fixed interspecific differences in 5S genes fail to accumulate, despite the fact that polymorphic nucleotide positions are common. **B** Diagrammatic representation of fitness curves as a function of the proportion of functional 5S genes in an array. Several possibilities are illustrated, although neither the actual shapes of these curves nor the level at which a selective threshold is reached is known. When the number of functional copies drops below a threshold level, selection can operate to remove arrays.

has preserved the 5S gene *consensus* sequence since the most recent common ancestor of the Malvaceae and Poaceae, or at least 120 million years.

Figure 5 presents a descriptive model of 5S rDNA evolution that incorporates the differential ability of gene and spacer sequences to fix novel variants (see also Schlötterer and Tautz 1994; Kellogg and Appels 1995). In this model concerted evolutionary forces eliminate variants or spread them throughout individual arrays, but not between different arrays, in keeping with our observations on allopolyploid species. Most nucleotide positions in the spacer region are presumed to be free to vary because variants are selectively neutral or near-neutral. Consequently, variant nucleotides can become either fixed or lost, thereby causing fixed interspecific differences to accumulate. In contrast, most mutations in the 5S gene are presumed to be selectively neutral or near-neutral *only when they occur in a subcritical proportion of repeats in an array.* Because 5S genes are present in high copy number, departures from the consensus sequence (perhaps even leading to nonfunctional 5S rRNAs) in a small proportion of genes are expected to have little overall effect on the fitness of an organism due to the buffering effect of functional 5S genes. Polymorphisms are therefore expected at "moderate" frequency. As variants move toward fixation by concerted evolution and stochastic factors, however, the number of functional 5S genes is reduced. Once this threshold frequency is reached (Fig. 5B), variant nucleotides become disadvantageous due to their effects on 5S transcription or 5S RNA function, the relative fitness of the entire *array* is thereby reduced (Williams 1990). As a consequence, fixed interspecific differences in 5S genes fail to accumulate, despite the fact that polymorphic nucleotide positions are as common as those in the spacer. In this respect, our results are consistent with those previously observed in 5S rDNA from diploid *Triticeae* (Kellogg and Appels 1995) and ITS sequences from *Drosophila* (Schlötterer and Tautz 1994), suggesting that this generalized model may have broad applicability.

The degree of 5S rDNA polymorphism observed in *Gossypium* and in other groups (Kellogg and Appels 1995) raises important questions concerning the biological consequences of heterogeneity in 5S genes. Because of low sequence similarity between plant 5S rDNA and homologues from model organisms such as *Xenopus, Drosophila,* and *Neurospora,* it is difficult to evaluate the effect of substitutions within previously defined transcription signals and control regions on rRNA transcription and/or function. At present, only two criteria, gross mutation/rearrangements in putative regulatory regions and an unexpectedly high accumulation of substitutions (as in putative pseudogenes such as *hirsutum6* and *hirsutum10*), may be used to assess the likelihood of 5S genes as being either functional or nonfunctional. If the 5S rDNA sequences reported here are representative of the genes that are transcribed, then the 5S rRNA pool within each species is heterogeneous. At present, no data address whether 5S rRNAs are as polymorphic as the genes that encode them. Further, it has yet to be determined whether the 5S rRNA pool is a random sample of the 5S rDNA genes or whether there are transcriptional consequences of genic and spacer polymorphisms. Finally, if only a subset of 5S repeats lead to functional 5S RNAs, what mechanisms promote selective transcription and/or filter out less-than-optimal rRNAs? Answers to these functional and mechanistic questions are essential to achieving a more complete understanding of 5S rDNA evolution.

Acknowledgments. We thank C. Brubaker, E. Kellogg, and A. Schnabel for comments on the manuscript; E. Kellogg for sharing unpublished data; Xiaoling Ding for assistance in automated sequencing; and the USDA Plant Genome Research Program and Texas Higher Education Coordinating Board (A.H.P.) and the National Science Foundation for financial support (J.F.W.).

References

Appels R, Baum BR, Clark BC (1992) The 5S DNA units of bread wheat (*Triticum aestivum*). Plant Syst Evol 183:183–194

Appels R, Honeycutt RL (1986) rDNA: evolution over a billion years. In: Dutta SK (ed) *DNA* systematics, vol. II. CRC Press, Boca Raton, FL, pp 81–155

Arnheim, N (1983) Concerted evolution of multigene families. In: Nei M, Koehn RK (eds) Evolution of genes and proteins. Sinauer, Sunderland, MA, pp 38–61

Basten CJ, Ohta T (1992) Simulation study of a multigene family, with special reference to the evolution of compensatory advantageous mutations. Genetics 132:247–252

Baum BR, Appels R (1992) Evolutionary change at the *5S Dna* loci of species in the Triticeae. Plant Syst Evol 183:195–208

Baum BR, Johnson DA (1994) The molecular diversity of the 5S rRNA gene in barley (*Hordeum vulgare*). Genome 37:992–998

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

Brubaker CL, Wendel JF (1993) On the specific status of *Gossypium lanceolatum* Todaro. Genet Resources Crop Evol 40:165–170

Campell BR, Song Y, Posch TE, Cullis CA, Town CD (1992) Sequence and organization of 5S ribosomal RNA-encoding genes of *Arabidopsis thaliana.* Gene 112:225–228

Capesius I (1991) Sequence of the 5S ribosomal RNA gene from *Sinapis alba.* Plant Mol Biol 17:169–170

Cox AV, Bennett MD, Dyer TA (1992) Use of the polymerase chain reaction to detect spacer size heterogeneity in plant 5S-rRNA gene clusters and to locate such clusters in wheat (*Triticum aestivum* L.). Theor Appl Genet 83:684–690

Crane CF, Price HJ, Stelly DM, Czeshin DG, McKnight TD (1993) Identification of a homeologous chromosome pair by in situ DNA hybridization to ribosomal RNA loci in meiotic chromosomes of cotton (*Gossypium hirsutum*). Genome 36:1015–1022

DeJoode DR (1992) Molecular insights into speciation in the genus *Gossypium* L. (Malvaceae). MS thesis, Iowa State University, Ames, IA

DeJoode DR, Wendel JF (1992) Genetic diversity and origin of the Hawaiian Islands cotton, *Gossypium tomentosum,* Am J Bot 79: 1311–1319

Devereux J, Haeberli P, Smithies O (1984) A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res 12: 387–395

Donoghue MJ, Olmstead RG, Smith JF, Palmer JD (1992) Phylogenetic relationships of Dipsacales based on *rbc*L sequences. Ann Mo Bot Garden 79:333–345

Dover GA (1982) Molecular drive: a cohesive mode of species evolution. Nature 299:111–117

Dover GA (1994) Concerted evolution, molecular drive and natural selection. Curr Biol 4:1165

Dvorák J, Zhang H-B, Kota RS, Lassner M (1989) Organization and evolution of the 5S ribosomal RNA gene family in wheat and related species. Genome 32:1003–1016

Dvorák J (1990) Evolution of multigene families: the ribosomal RNA loci of wheat and related species. In: Brown AHD, Clegg MT, Kahler AL, Weir BS (eds) Plant population genetics, breeding and genetic resources. Sinauer, Sunderland, MA, pp 83–97

Edwards GA, Endrizzi JE, Stein R (1974) Genome DNA content and chromosome organization in *Gossypium.* Chromosoma 47:309–326

Endrizzi JE, Turcotte EL, Kohel RJ (1985) Genetics, cytology, and evolution of *Gossypium.* Adv Genet 23:271–375

Fryxell PA (1979) *The natural history of the cotton tribe.* Texas A&M Univ Press, College Station, TX

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

Gerbi SA (1985) Evolution of ribosomal DNA. In: MacIntyre RJ (ed) Molecular evolutionary genetics. Plenum, NY, pp 419–490

Gottlob-McHugh SG, Lévesque M, MacKenzie K, Olson M, Yarosh O, Johnson DA (1990) Organization of the 5S rRNA genes in the soybean *Glycine max* (L). Merrill and conservation of the 5S rDNA repeat structure in higher plants. Genome 33:486–494

Halanych KM (1991) 5S Ribosomal RNA sequences inappropriate for phylogenetic reconstruction. Mol Biol Evol 8:249–253

Hemleben V, Werts D (1988) Sequence organization and putative regulatory elements in the 5S rRNA genes of two higher plants (*Vigna radiata* and *Matthiola incana*). Gene 62:165–169

Hood L, Campbell JH, Elgin SCR (1975) The organization, expression, and evolution of antibody genes and other multigene families. Annu Rev Genet 9:305–353

Kadir ZBZ (1976) DNA evolution in the genus *Gossypium.* Chromosoma 56:85–94

Kanazin V, Ananiev E, Blake T (1993) The genetics of 5S rRNA encoding multigene families in barley. Genome 36:1023–1028

Kellogg EA, Appels R (1995) Intraspecific and interspecific variation in 5S RNA genes are decoupled in diploid wheat relatives. Genetics 140:325–343

Kimura M (1980) A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J Mol Evol 16:111–120

Korn LJ (1982) Transcription of *Xenopus* 5S ribosomal RNA genes. Nature 295:101–105

Kumar S, Koichir T, Nei M (1993) MEGA, molecular evolutionary genetics analysis, v 1.0. Penn State Univ, University Park, PA

Li W-S, Luo C-C, Wu C-I (1985) Evolution of DNA Sequences. In: MacIntyre RJ (ed) Molecular evolutionary genetics. Plenum, NYC, NY, pp 1–94

Linares AR, Bowen T, Dover GA (1994) Aspects of nonrandom turnover involved in the concerted evolution of intergenic spacers within the ribosomal DNA of *Drosophila melanogaster*. J Mol Evol 39:151–159

Long EO, Dawid IB (1980) Repeated genes in eukaryotes. Annu Rev Biochem 49:727–764

Maddison DR (1991) The discovery and importance of multiple islands of most-parsimonious trees. Syst Zool 40:315–328

Masterson J (1994) Stomatal size in fossil plants: evidence for polyploidy in majority of angiosperms. Science 264:421–424

McDade LA (1990) Hybrids and phylogenetic systematics I. Patterns of character expression in hybrids and their implications for cladistic analysis. Evolution 44:1685–1700

McDade LA (1992) Hybrids and phylogenetic systematics II. The impact of hybrids on cladistic analysis. Evolution 46:1329–1346

McDonald JH, Kreitman M (1991) Adaptive protein evolution at the *Adh* locus in Drosophila. Nature 351:652–654

Michaelson MJ, Price HJ, Ellison JR, Johnston JS (1991) Comparison of plant DNA contents determined by Feulgen microspectrophotometry and laser flow cytometry. Am J Bot 78:183–188

Nagylaki T, Petes TD (1982) Intrachromosomal gene conversion and the maintenance of sequence homogeneity among repeated genes. Genetics 100:315–337

Nagylaki T (1984a) The evolution of multigene families under intrachromosomal gene conversion. Genetics 106:529–548

Nagylaki T (1984b) Evolution of multigene families under interchromosomal gene conversion. Proc Natl Acad Sci USA 81:3796–3800

Nagylaki T (1990) Gene conversion, linkage, and the evolution of repeated genes dispersed among multiple chromosomes. Genetics 126:261–276

Nei M (1987) Molecular evolutionary genetics. Columbia University Press, New York, NY

Ohta T, Dover GA (1983) Population genetics of multigene families that are dispersed into two or more chromosomes. Proc Natl Acad Sci USA 80:4079–4083

Ohta T (1983) On the evolution of multigene families. Theor Popul Biol 23:216–240

Ohta T (1984) Some models of gene conversion for treating the evolution of multigene families. Genetics 106:517–528

Ohta T (1990) How gene families evolve. Theor Popul Biol 37:213-219

Paterson AH, Brubaker CL, Wendel JF (1993) A rapid method for extraction of cotton (*Gossypium* spp.) genomic DNA suitable for RFLP or PCR analysis. Plant Mol Biol Rep 11:122–127

Percival AE (1987) The national collection of *Gossypium* germplasm. Southern Cooperative Series Bull 321, College Station, TX

Playford J, Appels R, Baum BR (1992) The 5S DNA units of *Acacia* species (Fabaceae). Plant Syst Evol 183:235–247

Rafalski JA, Wiewiorowski M, Söll D (1982) Organization and nucleotide sequence of nuclear 5S rRNA genes in yellow lupin (*Lupinus luteus*). Nucleic Acids Res 10:7635–7642

Reinisch AJ, Dong J, Brubaker CL, Stelly DM, Wendel JF, Paterson AH (1994) A detailed RFLP map of cotton, *Gossypium hirsutum* × *G. barbadense:* chromosome organization and evolution in a disomic polyploid genome. Genetics 138:829–847

Röder MS, Sorrells ME, Tanksley SD (1992) 5S ribosomal gene clusters in wheat: pulsed field gel electrophoresis reveals a high degree of polymorphism. Mol Gen Genet 232:215–220

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

Sambrook JE, Fritsch F, Maniatis T (1989) *Molecular cloning.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY

Sastri DC, Hilu K, Appels R, Lagudah ES, Playford J, Baum BR (1992) An overview of evolution in plant 5S DNA. Plant Syst Evol 183: 169–181

Schlötterer C, Tautz D (1994) Chromosomal homogeneity of *Drosophila* ribosomal DNA arrays suggests intrachromosomal exchanges drive concerted evolution. Curr Biol 4:777–783

Schneeberger RG, Creissen GP, Cullis CA (1989) Chromosomal and molecular analysis of 5S RNA gene organization in the flax, *Linum usitatissimum.* Gene 83:75–84

Scoles GJ, Gill BS, Xin Z-Y, Clarke BC, McIntyre CL, Chapman C, Appels R (1988) Frequent duplication and deletion events in the 5S RNA genes and the associated spacer regions of the Triticeae. Plant Syst Evol 160:105–122

Sharp SJ, Garcia AD (1988) Transcription of the *Drosophila melanogaster* 5S RNA gene requires an upstream promoter and four intragenic sequence elements. Mol Cell Biol 8:1266–1274

Smith GP (1976) Evolution of repeated DNA sequences by unequal crossover. Science 191:528–535

Sokal RR, Rohlf FJ (1981) Biometry. WH Freeman, San Francisco

Steele KP, Holsinger KE, Jansen RK, Taylor DW (1991) Assessing the reliability of 5S rRNA sequence data for phylogenetic analysis in green plants. Mol Biol Evol 8:240–248

Swofford DL (1990) PAUP: phylogenetic analysis using parsimony, version 3.1.1. Illinois Natural History Survey, Champaign, IL

Tyler BM (1987) Transcription of *Neurospora crassa* 5S rRNA genes requires a TATA box and three internal elements. J Mol Biol 196: 801–811

Vawter L, Brown WM (1993) Rates and patterns of base change in the small subunit ribosomal RNA gene. Genetics 134:597–608

VanderWiel PS, Voytas DF, Wendel JF (1993) *Copia*-like retrotransposable element evolution in diploid and polyploid cotton (*Gossypium* L.). J Mol Evol 36:429–447

Wendel JF (1989) New World tetraploid cottons contain Old World cytoplasm. Proc Natl Acad Sci USA 86:4132–4136

Wendel JF, Albert VA (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 JF, Percival AE (1990) Molecular divergence in the Galapagos Island–Baja California species pair, *Gossypium klotzschianum* and *G. davidsonii* (Malvaceae). Plant Syst Evol 171:99–115

Wendel JF, Olson PD, Stewart JM (1989) Genetic diversity, introgression and independent domestication of Old World cultivated cottons. Am J Bot 76:1795–1806

Wendel JF, Rowley R, Stewart J (1994) Genetic diversity in and phylogenetic relationships of the Brazilian endemic cotton, *Gossypium mustelinum* (Malvaceae). Plant Syst Evol 192:49–59

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

Wendel JF, Schnabel A, Seelanan T (1995b) An unusual ribosomal DNA sequence from *Gossypium gossypioides* reveals ancient, cryptic, intergenomic introgression. Mol Phyl Evol 4:298–313

Wheeler WC, Honeycutt RL (1988) Paired sequence difference in ribosomal RNAs: evolutionary and phylogenetic implications. Mol Biol Evol 5:90–96

White RJ (1994) RNA polymerase III transcription. RG Landes, Austin, TX

Williams S (1990) The opportunity for natural selection on multigene families. Genetics 124:439–441

Woese CR (1987) Bacterial evolution. Microbiol Rev 51:221–271

Wolters J, Erdmann VA (1988) Compilation of 5S rRNA and 5S rRNA gene sequences. Nucleic Acids Res 16(suppl):rl–r70

Zimmer EA, Martin SL, Beverley SM, Kan YW, Wilson Ac (1980) Rapid duplication and loss of genes coding for the alpha-chains of hemoglobin. Proc Natl Acad Sci USA 77:2158–2162

Zimmer EA, Jupe ER, Walbot VA (1988) Ribosomal gene structure, variation and inheritance in maize and its ancestors. Genetics 120: 1125–1136