American Journal of Botany 89(4): 707–725. 2002.

Rapid diversification of the cotton genus (*Gossypium*: Malvaceae) revealed by analysis of sixteen nuclear and chloroplast genes[1]

Richard C. Cronn,[2,4] Randall L. Small,[3] Tamara Haselkorn,[2] and Jonathan F. Wendel[2,5]

[2]Department of Botany, Iowa State University, Ames, Iowa 50011 USA; [3]Department of Botany, University of Tennessee, Knoxville, Tennessee 37996 USA

Previous molecular phylogenetic studies have failed to resolve the branching order among the major cotton (*Gossypium*) lineages, and it has been unclear whether this reflects actual history (rapid radiation) or sampling properties of the genes evaluated. In this paper, we reconsider the phylogenetic relationships of diploid cotton genome groups using DNA sequences from 11 single-copy nuclear loci (10 293 base pairs [bp]), nuclear ribosomal DNA (695 bp), and four chloroplast loci (7370 bp). Results from individual loci and combined nuclear and chloroplast DNA partitions reveal that the cotton genome groups radiated in rapid succession following the formation of the genus. Maximum likelihood analysis of nuclear synonymous sites shows that this radiation occurred within a time span equivalent to 17% of the time since the separation of *Gossypium* from its nearest extant relatives in the genera *Kokia* and *Gossypioides*. Chloroplast and nuclear phylogenies differ significantly with respect to resolution of the basal divergence in the genus and to interrelationships among African cottons. This incongruence is due to limited character evolution in cpDNA and either previously unsuspected hybridization or unreliable phylogenetic performance of the cpDNA characters. This study highlights the necessity of using multiple, independent data sets for resolving phylogenetic relationships of rapidly diverged lineages.

Key words: chloroplast DNA; *Gossypium*; nuclear DNA; nuclear markers; phylogenetic incongruence.

One of the more vexing problems in phylogenetic analysis is the occurrence of short interior branches. Because short internodes often have weak support as measured by jackknife and bootstrap resampling, decay analysis, or other indicators of relative confidence, subsequent phylogenetic analyses using additional molecular markers often fail to yield the same topology. Arising from long branch attraction, insufficient signal, or other causes, short internodes are frequently observed in phylogenetic trees and are thought to be a common cause of misleading phylogenetic inference as well as topological incongruence among data sets (reviewed in Wendel and Doyle, 1998). The short internode problem is a relative concept, dependent both on the scale of temporal divergence and the relative rates of character change, but it is applicable to all taxonomic levels in both plants and animals (e.g., Olmstead and Sweere, 1994; Fehrer, 1996; Lara, Patton, and DaSilva, 1996; Baldwin, 1997; Brinkmann and Philippe, 1999; Hughes and Baker, 1999; Kennedy et al., 1999; Waits et al., 1999). A common response when confronted with poorly supported or conflicting short interior branches is to sample more taxa to break up terminal branches, more characters to increase true signal, or both (Hillis, Huelsenbeck, and Cunningham, 1994; Cummings, Otto, and Wakeley, 1995; Hillis, 1996, 1998; but see Naylor and Brown, 1998; Nei, Kumar, and Takahashi, 1998; Poe, 1998; Soltis et al., 1998; Bremer et al., 1999; Poe and Swofford, 1999). Because phylogenetically inferred short interior branches may reflect relatively rapid radiations on a temporal scale that is compressed relative to the rate of character-state change, it may be exceedingly difficult, even impossible, to confidently resolve particular branch orders.

A probable example of rapid and global diversification is offered by the cotton genus *Gossypium* L. (Malvaceae). This genus of about 50 species (Fryxell, 1979, 1992) is differentiated cytogenetically into eight ‘‘genome groups’’ (designated ‘‘A’’ through ‘‘G’’, and ‘‘K’’) that differ in DNA content and chromosome size but not in chromosome number (reviewed in Endrizzi, Turcotte, and Kohel, 1985; Percival, Stewart, and Wendel, 1999; Wendel et al., 1999). The inclusivity of each genome group is congruent with the alignment of species in the most recent taxonomic treatment (Fryxell, 1979, 1992), in which three subgenera of diploid species are recognized: the African/Arabian subgenus *Gossypium*, containing A-, B-, E-, and F-genome cottons; the Australian subgenus *Sturtia* (R. Brown) Todaro (C-, G-, and K-genomes); and subgenus *Houzingenia* Fryxell, comprising the New World, D-genome cottons.

Present molecular evidence indicates that each genome group is monophyletic (Wendel and Albert, 1992; Seelanan, Schnabel, and Wendel, 1997; Seelanan et al., 1999), but relationships among genome groups remain in question (Wendel and Albert, 1992; Seelanan, Schnabel, and Wendel, 1997; Seelanan et al., 1999). As a consequence, branching orders among the three subgenera are unresolved, as is monophyly for the African/Arabian subgenus *Gossypium*. Using cpDNA restriction site data, Wendel and Albert (1992) suggested that the Australian cotton lineage (C-, G-, and K-genomes) represents the earliest divergence within the genus (Fig. 1). This interpretation was later shown to be based on modest character support (Seelanan, Schnabel, and Wendel, 1997); moreover, sequence data for the plastid gene *ndhF* failed to provide corroborating evidence of this relationship (Fig. 1). Uncertainty regarding relationships among genome groups is also apparent in the cpDNA restriction site and *ndhF* sequence analysis,

[1] Manuscript received 3 May 2001; revision accepted 20 September 2001.

We thank Q. Liu and C. Brubaker for sharing *FAD2-1* data prior to publication, M. Fishbein for helpful discussion regarding incongruence tests, and the National Science Foundation for support.

[4] Current address: Pacific Northwest Research Station, USDA Forest Service, 3200 SW Jefferson Way, Corvallis, Oregon 97331 USA.

[5] Author for reprint requests (e-mail: jfw@iastate.edu).

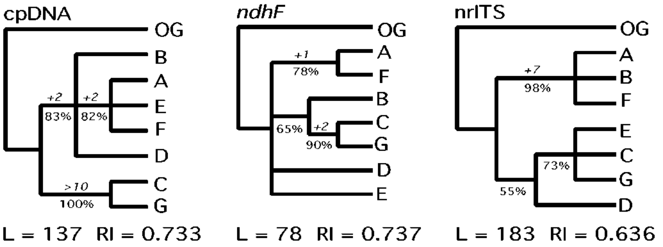


Fig. 1. Prior phylogenetic hypotheses for diploid *Gossypium* genome groups derived from cpDNA restriction sites (Wendel and Albert, 1992), chloroplast *ndhF* gene sequences, and nuclear ribosomal nrITS sequences (Seelanan, Schnabel, and Wendel, 1997). Illustrated are the strict consensus trees for four (cpDNA), six (*ndh*F), and four (nrITS) most parsimonious trees recovered from exhaustive searches. Decay values (*d*, italicized) and jackknife percentages (*j*) are shown for nodes where $d > 0$ and $j \geq 50\%$, and the treelength (L) and retention index (RI) is shown below each tree. *Thespesia populnea* was used as the outgroup (OG) for all analyses. Note the phylogenetic alternatives for the primary cladogenic event, as well as the equivocal resolution of the African B- and E-genome clades.

most notably in the equivocal placement of the African B-genome lineage (Wendel and Albert, 1992). This result raised doubts as to the monophyly of subgenus *Gossypium*, as have later studies based on phylogenetic analyses of nuclear rDNA sequences (Seelanan, Schnabel, and Wendel, 1997). Sequence data from the internal transcribed spacer region of the nuclear ribosomal internal transcribed spacer (nrITS) suggested that Arabian E-genome and Australian C-genome clades are sister groups, with the (E + C)-genome clade sister to the New World D-genome clade. This assemblage resolves as sister to the remaining African genome groups (Fig. 1; Seelanan, Schnabel, and Wendel, 1997). Thus, existing data sets are contradictory in two major aspects, first with regard to the monophyly of African/Arabian cottons and second with regard to the earliest divergence event in the genus.

In each of the foregoing examples, inspection revealed short internal branches for the conflicting resolutions. This incongruence may reflect intrinsic evolutionary properties of the genes used in these analyses, or it may instead reflect an important feature of *Gossypium* evolution, namely, a relatively rapid diversification of the genus after its origin (Seelanan, Schnabel, and Wendel, 1997; Wendel and Doyle, 1998). Our purpose in the present paper is to explore these possibilities using multiple data sets derived from the nuclear and plastid genomes. If the major clades within the genus originated in a temporally brief time period, we expect to consistently recover short interior branches, and possibly additional incongruence among newly generated gene trees, as a consequence of spurious relationships inferred from cases in which homoplasy overwhelms true signal.

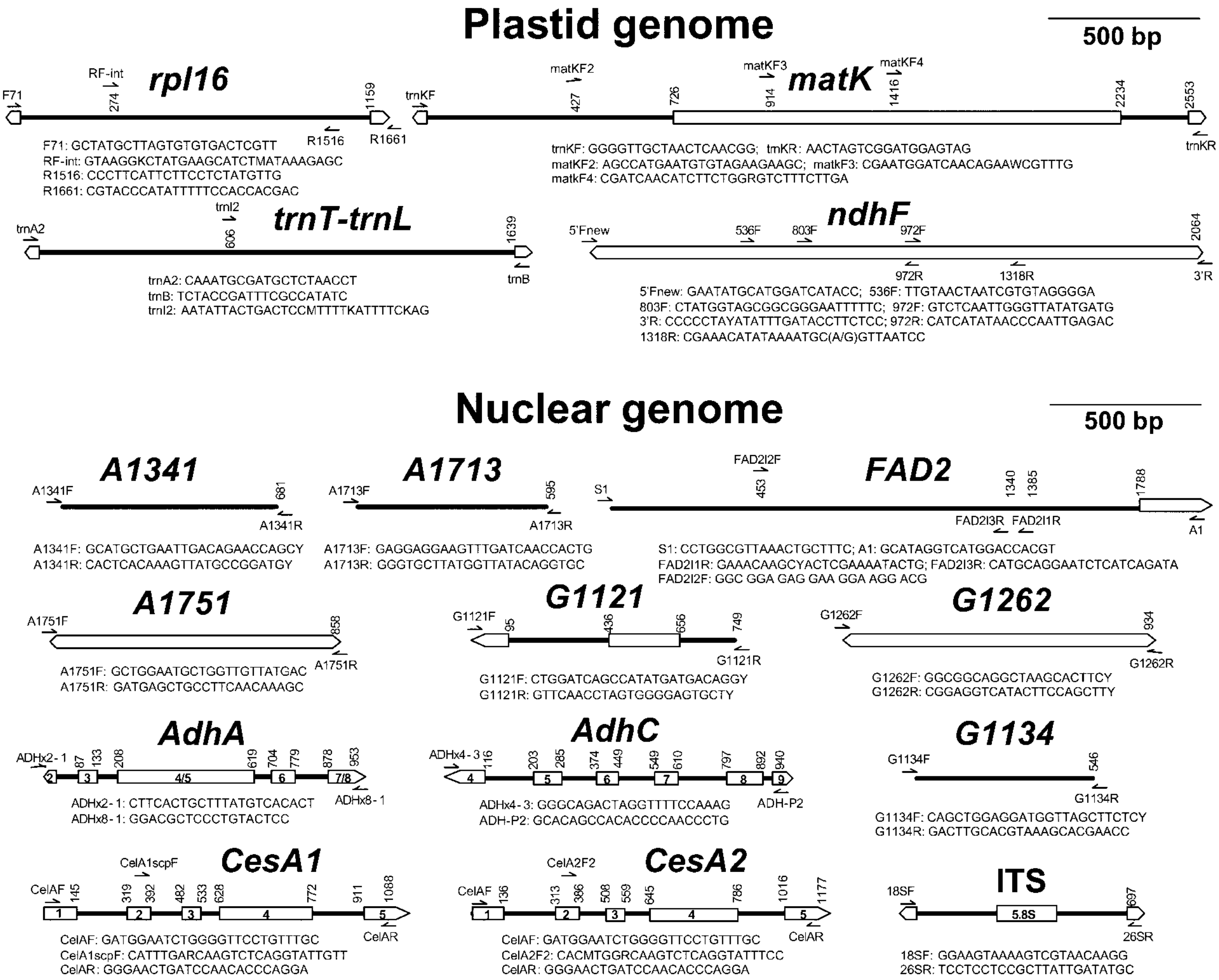
Our second objective was to elucidate the actual cladistic affinities of the major diploid cotton lineages, notwithstanding the challenges imposed by the suspected history of rapid radiation. Toward this end, we employed two rational approaches to resolving poorly supported nodes, namely, we used multiple genes and generated larger data sets (e.g., Hillis, Huelsenbeck, and Cunningham, 1994; Olmstead and Sweere, 1994; Cummings, Otto, and Wakeley, 1995; Hillis, 1996; Nei, Kumar, and Takahashi, 1998; Soltis et al., 1998; Bremer et al., 1999; Hughes and Baker, 1999; Walsh et al., 1999). In this paper, we reevaluate the issue of the cotton genomes phylogeny using three additional chloroplast DNA regions and 11 single-copy nuclear genes. By combining these data with preexisting chloroplast *ndhF* and nuclear nrITS sequence information (Seelanan, Schnabel, and Wendel, 1997), we were able to evaluate the phylogenetic signal contained within 18 kilobases (kb) of sequence data derived from the chloroplast (7 kb) and nuclear (11 kb) genomes. This substantial data set bolsters the relatively weak signal present in short interior branches and provides complete resolution of the branching order among major cotton lineages. The resolutions provided by both chloroplast and nuclear DNA enable comparisons among individual loci and data partitions, permit a multilocus evaluation of interior node resolution, and provide a detailed picture of the relative timing of cladogenic events that have given rise to extant diploid cotton genome groups.

MATERIALS AND METHODS

Plant materials—Taxa included in this study are listed in Table 1, and GenBank accession numbers for each taxon–locus combination have been archived under Appendix 1 at the Botanical Society of America website (http://ajbsupp.botany.org/v89/Cronn/.doc). Since the monophyly of *Gossypium* genome groups has been convincingly established (Wendel and Albert, 1992; Seelanan, Schnabel, and Wendel, 1997; Liu et al., 2001), we sampled one representative species from each: these include *G. herbaceum* (A-genome) *G. anomalum* (B-genome), *G. robinsonii* (C-genome), *G. raimondii* (D-genome), *G. somalense* (E-genome), *G. longicalyx* (F-genome), and *G. bickii* (G-genome). In three cases, we substituted placeholders for these genomes: *G. stocksii* and *G. arboreum* represented the E- and A-genomes, respectively, for *ndhF* analysis, and the A-subgenome sequence from the AD-genome allotetraploid *G. hirsutum* was used to represent the A-genome in *AdhC* analysis (Small et al., 1998). The distinctive K-genome cottons from northwest Aus-

TABLE 1. Taxa included in the present study.

Subgenus	Taxon	Genome	Accession (voucher)	Geographic origin
Gossypium	*G. herbaceum* L. subsp. *africanum* (Watt) Mauer	A_1	A1-73 (JFW539)	Botswana
	G. arboreum L.	A_2	A2-47 (JFW & TDC305)	Sudan
	G. anomalum Wawra ex Wawra & Peyritch	B_1	B1-1 (JFW & TDC 312)	Africa
	G. triphyllum (Harvey & Sonder) Hochreutiner	B_2	(TS16)	SW Africa
	G. somalense (Gurke) J. B. Hutchinson	E_2	(TS14)	NE Africa
	G. stocksii Masters	E_1	(TS13)	NE Africa
	G. longicalyx J. B. Hutchinson & Lee	F_1	F1-1 (TS8)	Tanzania
Sturtia	*G. robinsonii* F. von Mueller	C_2	(TS12)	Australia
	G. bickii Prokhanov	G_1	(JFW & TDC557)	Australia
Houzingenia	*G. raimondii* Ulbrich	D_5	''Galau's'' (JFW & TDC591)	Peru
Karpas	*G. hirsutum* L.	AD_1	race Palmeri (JFW & TDC632)	Oaxaca, Mexico
Outgroup 1	*Kokia drynarioides* (Seemann) Lewton	—	(TS6)	Hawaiian Islands
Outgroup 2	*Gossypioides kirkii* (Mast.) J. B. Hutchinson	—	(TS3)	E. Africa, Madagascar



F71: GCTATGCTTAGTGTGTGACTCGTT
RF-int: GTAAGGKCTATGAAGCATCTMATAAAGAGC
R1516: CCCTTCATTCTTCCTCTATGTTG
R1661: CGTACCCATATTTTTCCACCACGAC

trnKF: GGGGTTGCTAACTCAACGG; trnKR: AACTAGTCGGATGGAGTAG
matKF2: AGCCATGAATGTGTAGAAGAAGC; matKF3: CGAATGGATCAACAGAAWCGTTTG
matkF4: CGATCAACATCTTCTGGRGTCTTTCTTGA

trnA2: CAAATGCGATGCTCTAACCT
trnB: TCTACCGATTTCGCCATATC
trnI2: AATATTACTGACTCCMTTTTKATTTTCKAG

5'Fnew: GAATATGCATGGATCATACC; 536F: TTGTAACTAATCGTGTAGGGGA
803F: CTATGGTAGCGGCGGGAATTTTTC; 972F: GTCTCAATTGGGTTATATGATG
3'R: CCCCCTAYATATTTGATACCTTCTCC; 972R: CATCATATAACCCAATTGAGAC
1318R: CGAAACATATAAAATGC(A/G)GTTAATCC

A1341F: GCATGCTGAATTGACAGAACCAGCY
A1341R: CACTCACAAAGTTATGCCGGATGY

A1713F: GAGGAGGAAGTTTGATCAACCACTG
A1713R: GGGTGCTTATGGTTATACAGGTGC

S1: CCTGGCGTTAAACTGCTTTC; A1: GCATAGGTCATGGACCACGT
FAD2I1R: GAAACAAGCYACTCGAAAATACTG; FAD2I3R: CATGCAGGAATCTCATCAGATA
FAD2I2F: GGC GGA GAG GAA GGA AGG ACG

A1751F: GCTGGAATGCTGGTTGTTATGAC
A1751R: GATGAGCTGCCTTCAACAAAGC

G1121F: CTGGATCAGCCATATGATGACAGGY
G1121R: GTTCAACCTAGTGGGGAGTGCTY

G1262F: GGCGGCAGGCTAAGCACTTCY
G1262R: CGGAGGTCATACTTCCAGCTTY

ADHx2-1: CTTCACTGCTTTATGTCACACT
ADHx8-1: GGACGCTCCCTGTACTCC

ADHx4-3: GGGCAGACTAGGTTTTCCAAAG
ADH-P2: GCACAGCCACACCCCAACCCTG

G1134F: CAGCTGGAGGATGGTTAGCTTCTCY
G1134R: GACTTGCACGTAAAGCACGAACC

CelAF: GATGGAATCTGGGGTTCCTGTTTGC
CelA1scpF: CATTTGARCAAGTCTCAGGTATTGTT
CelAR: GGGAACTGATCCAACACCCAGGA

CelAF: GATGGAATCTGGGGTTCCTGTTTGC
CelA2F2: CACMTGGRCAAGTCTCAGGTATTTCC
CelAR: GGGAACTGATCCAACACCCAGGA

18SF: GGAAGTAAAAGTCGTAACAAGG
26SR: TCCTCCTCCGCTTATTGATATGC

Fig. 2. Gene diagrams for the 12 nuclear and 4 chloroplast loci included in this study. Gene diagrams show the overall length sampled from each locus, as well the location of exon (shaded bars) and intron or noncoding (black line) regions. Primers used for amplification and sequencing are indicated by arrows either above (forward primers) or below (reverse primers) each gene diagram; the name and nucleotide sequence for all primers used are indicated below gene diagrams.

tralia were omitted from this study because all phylogenetic analyses to date place K-genome species in a monophyletic clade allied with other Australian cottons (Wendel and Albert, 1992; Seelanan, Schnabel, and Wendel, 1997; Seelanan et al., 1999; Liu et al., 2001). We chose *Gossypioides kirkii* and *Kokia drynarioides* as outgroup taxa because these two genera comprise the sister lineage to *Gossypium* (Seelanan, Schnabel, and Wendel, 1997). Leaf DNA from each species was extracted using the method of Paterson, Brubaker, and Wendel (1993). Voucher specimens were deposited in the Hayden Herbarium (ISC) at Iowa State University, Ames, Iowa, USA.

Chloroplast loci examined—Three loci from the plastid genome were amplified using the polymerase chain reaction (PCR), isolated and sequenced from all ingroup and outgroup taxa: the *matK* gene and flanking *trnK* intron sequences (Steele and Vilgalys, 1994), the *rpl16* intron (Jordan, Courtney, and Neigel, 1996), and the *trnT-trnL* spacer (Taberlet et al., 1991). Primers used to amplify these loci follow the original references and are summarized in Fig. 2. Amplification conditions for each locus followed previously optimized protocols (Small et al., 1998). Following amplification, reaction products were gel-purified and directly sequenced using the ABI ''Big Dye'' dideoxynucleotide termination kit (Applied Biosystems, Foster City, California, USA). All reactions used approximately 100 ng of PCR product and 5 pmol of sequencing primer. Sequences were resolved on an ABI377 (Applied Biosystems) automated sequencer at the Iowa State University DNA Sequencing and Synthesis Facility, and automated traces were evaluated using Sequence Navigator (version 1.0.1, Applied Biosystems). In addition to the cpDNA regions noted above, we included *ndhF* sequences reported previously (Seelanan, Schnabel, and Wendel, 1997).

Nuclear loci examined—To evaluate phylogenetic signal from the nuclear genome, we generated gene trees for 12 nuclear loci (11 reported here, plus the nrITS sequences from Seelanan, Schnabel, and Wendel, 1997). Gene diagrams showing primer sequences and intron/exon structures are shown in Fig. 2. All loci were amplified, isolated, and sequenced from all ingroup and outgroup taxa using primers and conditions described in Cronn, Small, and Wendel (1999). These loci include six anonymous markers that correspond to *PstI* mapping probes *A1341, A1713, A1751, G1121, G1134,* and *G1262* (Brubaker and Wendel, 1994; Cronn and Wendel, 1998; Cronn, Small, and Wendel, 1999), the two alcohol dehydrogenase genes *AdhA* and *AdhC* (Small et al., 1998; Cronn, Small, and Wendel, 1999; Small and Wendel, 2000), two cellulose synthase genes (*CesA1* and *CesA1b,* formerly *CelA1* and *CelA2*; Cronn, Small, and Wendel, 1999; R. Cronn et al., unpublished data) and a 5′ untranslated intron from a microsomal fatty acid desaturase gene (*FAD2-1*; Liu et al., 2001). An important criterion in selecting the nuclear loci employed in this study is that, with the exception of rDNA, all genes have been shown to be low-copy or unique in A- and D-genome diploid cottons by Southern hybridization experiments (Brubaker and Wendel, 1994; Cronn and Wendel, 1998; Brubaker, Paterson, and Wendel, 1999; Small and Wendel, 2000; R. Cronn et al., unpublished data). Additionally, orthological relationships for all markers (except *FAD2-1*) have been established by comparative genetic mapping, whereby the placement of putative orthologues were compared on A-genome and D-genome linkage maps that include over 300 molecular markers (Cronn and Wendel, 1998; Brubaker, Paterson, and Wendel, 1999; Small and Wendel, 2000; R. Cronn and R. Small, unpublished data). The combination of low-copy number and localization of all genes to syntenic, colinear linkage groups in the diploid A- and D-genomes constitutes rigorous evidence of orthology.

Amplification primers and protocols used to amplify each of these loci follow previously detailed protocols (Small et al., 1998; Cronn, Small, and Wendel, 1999; Liu et al., 2001). Polymerase chain reaction products were gel-purified and directly sequenced for all loci except the *FAD2-1* intron. Due to apparent polymerase slippage in the numerous stretches of AT-rich sequence, we found it necessary to clone *FAD2-1* PCR products (using pGem-T Easy; Promega, Madison, Wisconsin, USA) prior to sequencing. Sequencing reactions for cloned *FAD2-1* inserts used 500 ng of plasmid and 5 pmol of sequencing primer. The consensus sequence of three clones was determined from each taxon to reduce the errors induced by *Taq* polymerase.

Data analysis—Sequences were aligned with the Clustal function of Sequence Navigator using default parameters. Sequence alignment led to the recognition of numerous gaps in both the individual and combined data sets. While many of these gaps could be aligned with high confidence in assessments of positional homology, five regions from four genes showed complex, possibly superimposed indel patterns. These regions, which consequently were excluded from all analyses, include positions 473–521 of the chloroplast *rpl16* intron (2544–2587 of the combined chloroplast data set); positions 294–450 of the chloroplast *trnT-trnL* spacer (3516–3672 of the combined chloroplast data set); positions 2510–2552 of the *trnK* intron downstream from the chloroplast *matK* gene (7328–7370 of the combined chloroplast data set); and positions 935–1015 and 1529–1587 of the nuclear *FAD2-1* locus (10 150–10 230 and 10 744–10 802 of the combined nuclear data set). The remaining alignment gaps were treated as missing data in all analyses. To evaluate the phylogenetic signal contributed by insertion/deletion (indel) events, we scored all indels in the cpDNA and nDNA data sets for which we had confidence in positional homology (archived as Appendix 2 at the Botanical Society of America website, http://ajbsupp.botany.org/v89/Cronn/.doc) and analyzed these data separately from nucleotide data using parsimony analysis. Indels were coded as unordered characters with binary states (in the case of simple presence/absence indels) or multistate characters (in the case of indels with variable length but one identical 5′ or 3′ end) using the criteria described in Simmons and Ochoterena (2000). Due to their high lability, di- and trinucleotide repeats found in chloroplast *rpl16* intron (TA$_{2\text{-}4}$), chloroplast *trnK* (*matK*) intron (TTA$_{2\text{-}4}$), chloroplast *trnT-trnL* intron (GAA$_{3\text{-}5}$), and nuclear *FAD2-1* intron (GAA$_{4\text{-}10}$) data sets were specifically excluded from this analysis. Complete sequence data for all loci are available from the author at http://www.botany.iastate.edu/~jfw/HomePage/jfwdata_sets.html or from the Botanical Society of America website (http://www.botany.org/bsa/ajbsupp/).

Phylogenetic analyses, based on maximum parsimony (MP) and maximum likelihood (ML) methods, were performed using PAUP* version 4.0b8 (Swofford, 2001). Most-parsimonious trees were found from exhaustive searches with equal weighting of all characters. Maximum likelihood trees were estimated using the Hasegawa-Kishino-Yano (1985) model of sequence evolution, with heuristic searches and tree bisection-reconstruction (TBR) branch swapping. For each analysis, we used empirical base frequencies and estimated the following parameters: the transition/transversion rate, the gamma shape parameter (using four rate categories), and the proportion of invariable sites. We explored other models of sequence evolution and combinations of input parameters but found that these perturbations had no effect on tree topologies and did not significantly improve the likelihood scores.

To evaluate ‘‘combinability’’ (Bull et al., 1993; Johnson and Soltis, 1998) of the 16 loci included in this study, incongruence length difference tests (ILD; Farris et al., 1995) were performed pairwise on all loci within a genome partition (either ‘‘chloroplast’’ or ‘‘nuclear’’), on two combined data sets representing all sequences from either the chloroplast (cpDNA) or nuclear (nDNA) genome, and on the ‘‘indel’’ data sets derived from cpDNA and nDNA. These tests were implemented using PAUP* 4.0 (as the partition-homogeneity test). Each analysis utilized 1000 replicates, in which random partitions of equal size were created by sampling sites without replacement from the original data. Replicates were analyzed using the parsimony criterion, and branch swapping using TBR was performed on trees held at each step during the stepwise addition.

Several methods were used to evaluate support of tree topologies derived from individual and combined data sets. For parsimony analyses, jackknife resampling (Farris, 1996) was performed using 1000 replicates (branch and bound searches) and with 50% of the characters deleted in each replicate. Decay analysis (Bremer, 1988) was performed by searching exhaustively for all trees up to ten steps longer than the most-parsimonious and noting how many steps longer than the shortest tree each clade survived. For ML analyses, the likelihood ratio test of Kishino and Hasegawa (KH test; 1989) was performed on the combined nuclear and chloroplast data sets to test among alternative topologies obtained from prior phylogenetic studies (detailed in *RESULTS*).

To estimate the relative divergence times of the major diploid cotton lineages, ML methods were used to evaluate clocklike rate constancy across lineages and to calculate likelihood-based branch lengths. In cases of rate constancy, branch lengths from the optimal ML topology are directly proportional to time (Baum, Small, and Wendel, 1998; Sanderson, 1998). To test the assumption of clocklike sequence divergence, the likelihoods of clock-enforced and clock-unenforced topologies were computed (using the parameters described above) from reduced data sets composed of nuclear DNA synonymous sites (7978 aligned sites). Since clock-enforced and clock-unenforced analyses yielded identical, single trees from the nuclear DNA data set, we used the likelihood-ratio test (Sanderson, 1998) to evaluate whether the clock-enforced likelihood score differed significantly from the clock-unenforced alternative. To convert relative divergence time into illustrative absolute divergence dates, we first computed pairwise synonymous divergences (K_S, using Jukes-Cantor transformation) for silent sites in exons and introns between all taxa, using the computer program DnaSP version 3 (Rozas and Rozas, 1999). We then computed the mean K_S between all ingroup taxa ($N = 7$) and the two outgroup taxa. This value ($K_S = 0.0672 \pm 0.0556$ substitutions per site; $N = 14$ comparisons) was converted into absolute time using the following equation: $T_{\text{divergence}} = K_S\ /(2)(r)$, where r corresponds to the absolute rate of synonymous site divergence in nuclear *Adh* genes from palms (2.6×10^{-9} substitutions · synonymous site^{-1} · yr^{-1}; Morton, Gaut, and Clegg, 1996) or members of the Brassicaceae (1.5×10^{-8} substitutions · synonymous site^{-1} · yr^{-1}; Koch, Haubold, and Mitchell-Olds, 2000.) Although there are numerous potential sources for error in molecular clock estimates of divergence (Sanderson, 1998), estimates of absolute divergence times are presented for heuristic purposes, as well as to compare to other divergence time estimates for *Gossypium* (Endrizzi, Katterman, and Geever, 1989; Geever, Katterman, and Endrizzi, 1989; Wendel and Albert, 1992; Seelanan, Schnabel, and Wendel, 1997).

RESULTS

Sequence data were generated for 14 different genes and spacer regions, yielding final data sets that include 4 chloroplast (combined length = 7370 bp) and 12 nuclear loci (combined length = 10 988 bp). Results from two of the genes included in the present study, chloroplast *ndhF* and nuclear nrITS, have already previously been described in the context of evaluating the phylogeny of the *Gossypieae* (Seelanan, Schnabel, and Wendel, 1997). These two data sets were included in the present analysis for comparative analyses of relative rates of sequence evolution and so that the conflicting phylogenies derived from *ndhF* and nrITS sequence data could be evaluated with respect to the other loci reported here. Due to space limitations, we summarize descriptive statistics for the 16 data sets in Table 2. This table highlights the most

TABLE 2. Description of regions sequenced from the plastid (top) and nuclear (bottom) genomes.[a]

Locus	Aligned length	Length range	Exon length	Intron bp, aligned	Intron bp, range	PI indels	Sites excluded	Variable sites (%)	PI sites (%)	Ratio of PI : variable sites
Plastid genome										
ndhF	2064	2041–2064	2064	0	0	1	0	47 (2.27)	11 (0.54)	0.234
matK	2553	2474–2536	1509	1044	955–1027	7	42	62 (2.47)	20 (0.77)	0.323
rpl16	1157	1080–1135	0	1157	1080–1135		51	22 (1.99)	4 (0.36)	0.182
trnT-trnL	1596	1138–1421	0	1596	1138–1421		156	77 (5.34)	14 (0.97)	0.182
Totals	7370		3573	3797		24	249	208 (2.92)	49 (0.69)	0.236
Nuclear genome										
A1341	681	663–670	0	681	663–670	2	28	47 (7.20)	4 (0.61)	0.085
A1713	595	580–593	0	594	580–593	0	20	34 (5.91)	3 (0.52)	0.088
A1751	858	858	858	0	0	0	47	28 (3.45)	2 (0.27)	0.071
AdhA	953	941–950	617	336	323–333	3	19	50 (5.35)	4 (0.43)	0.080
AdhC	940	918–930	433	506	485–497	8	60	61 (6.85)	7 (0.79)	0.115
CesA1	1086	1072–1086	594	493	479–493	2	152	65 (6.10)	4 (0.38)	0.061
CesA1b	1177	1116–1175	566	611	550–609	3	75	103 (9.25)	5 (0.45)	0.048
FAD2-1	1774	1235–1630	0	1774	1235–1630	10	751	124 (12.1)	17 (1.66)	0.137
G1121	748	719–748	316	431	403–433	2	34	30 (4.20)	4 (0.56)	0.133
G1134	547	546	0	546	546	0	0	24 (4.40)	5 (0.92)	0.208
G1262	934	888–934	934	0	0	0	0	32 (3.43)	2 (0.21)	0.063
nrITS	695	668–688	0	696	668–688	3	16	106 (15.6)	28 (4.12)	0.264
Totals	10988		4318	6668		33	1202	704 (6.89)	85 (0.83)	0.121

[a] Variable and phylogenetically informative (PI) sites are for ingroup comparisons; excluded sites are those removed from the data set prior to phylogenetic analysis due to alignment gaps or alignment ambiguity. Lengths are in base pairs (bp).

important features of each data set and includes absolute and aligned length, number of exon and intron nucleotides, the number of variable and phylogenetically informative sites, and the number of phylogenetically informative indels in each data set.

Characteristics of combined data sets—*Combined chloroplast sequences*—Aligned sequences from *ndhF* (2064 bp), *matK* (2553 bp), the rpl16 intron (1157 bp), and the trnT-trnL spacer (1596 bp) were concatenated into a single data set (the ‘‘cpDNA’’ data set). Incongruence-length difference (ILD) tests on all pairwise combinations of the four individual chloroplast sequences yielded no significant incongruence (P = 0.20–1.00); accordingly, we combined the sequences into a concatenated data set. Concatenated sequences ranged in length from 6875 bp (*Kokia drynarioides*) to 7109 bp (*Gossypioides kirkii*). Alignment of concatenated sequences yielded a data set 7370 bp in length, consisting of 3573 exon nucleotides and 3797 intron or spacer nucleotides. Of the 7370 aligned nucleotide positions, 249 positions were excluded due to alignment uncertainty (51 bp from *rpl16* intron, 156 bp from *trnT-trnL* spacer, 42 bp from *matK*), yielding a final data set of 7121 bp, with a nucleotide composition of 16.8% G, 16.2% C, 29.1% A, and 37.9% T. A total of 54 fixed substitutions separated the ingroup and outgroup, 208 sites (2.9%) were variable, and 49 sites (0.7%) were phylogenetically informative within the ingroup. Pairwise Jukes-Cantor distances ranged from 0.0205 between *G. robinsonii* and *Gossypioides kirkii* to 0.0048 between the A- and F-genomes (Table 3). The average Jukes-Cantor distance among ingroup taxa was 0.0103 ± 0.0024 (mean ± standard deviation; N = 21), and the average distance between all ingroup taxa and the two outgroup taxa was 0.0182 ± 0.0017.

Combined nuclear sequences—Aligned sequences from *A1341* (681 bp), *A1713* (595 bp), *A1751* (858 bp), *AdhA* (953 bp), *AdhC* (940 bp), *CesA1* (1086 bp), *CesA1b* (1177 bp), *FAD2-1 intron* (1774 bp), *G1121* (748 bp), *G1134* (547 bp), *G1262* (934 bp), and nrITS (695 bp) were concatenated into a single data set (the ‘‘nDNA’’ data set). Pairwise ILD tests on all individual nuclear sequences (66 possible pairwise tests) yielded only a single instance of statistical conflict, namely, between the *FAD2-1* intron and nrITS data sets (P = 0.001). These two data sets exhibited the greatest phylogenetic signal among nuclear genes, so their pattern heterogeneity likely reflects different evolutionary constraints or histories. In this regard, we note that nrITS from diploid and polyploid *Gossypium* have previously been shown to exhibit unexpected patterns of sequence evolution, including non-Mendelian inheritance in allopolyploids (Wendel, Schnabel, and Seelanan, 1995a) and apparent sequence chimera formation in suspected diploid hybrids (Wendel, Schnabel, and Seelanan, 1995b). Despite this single result of pattern heterogeneity, comparisons between *FAD2-1* intron and nrITS and the remaining ten nuclear genes exhibited no additional heterogeneity. For this reason, we chose to analyze nuclear data sets sequences individually, as well as combined. The concatenated sequences ranged in length from 10 330 bp (in *G. raimondii*) to 10 839 bp (*Kokia drynarioides*). Alignment of these sequences yielded a data set that was 10 988 bp in length, consisting of 4318 bp of exon (39.2%) and 6668 bp of intron or spacer DNA (60.8%). This data set provided a cumulative total of 33 phylogenetically informative gaps. Of 10 988 bp, 138 positions (10 150–10 230 and 10 744–10 802 in the combined alignment) were excluded due to alignment uncertainties in the *FAD2-1* intron data set. The final nuclear DNA data set used for phylogenetic analysis included 10 850 bp and showed a nucleotide composition of 20.8% G, 18.2% C, 28.9% A, and 32.1% T. A total of 226 fixed substitutions separated the ingroup from outgroup. The data matrix included 704 sites (6.89%) that were variable within the ingroup, and of these, 85 (0.83%) were phylogenetically informative. Comparisons among the composite sequences yielded pairwise nucleotide distances that ranged from a high of 0.0557 between *G. so-*

TABLE 3. Summary of Jukes-Cantor distances[a] for all sites (bold typeface), silent sites[b] (italic typeface), and replacement sites (plain typeface) in pairwise comparisons of the cpDNA data set (four chloroplast genes; above diagonal) and the nDNA data set (12 nuclear loci; below diagonal).

Taxon	A	B	C	D	E	F	G	*Gk*[c]	*Kd*[c]
A	—	**0.0099**	**0.0133**	**0.0108**	**0.0099**	**0.0048**	**0.0126**	**0.0190**	**0.0168**
		0.0129	*0.0158*	*0.0128*	*0.0119*	*0.0067*	*0.0153*	*0.0241*	*0.0198*
		0.0051	0.0084	0.0084	0.0055	0.0015	0.0070	0.0092	0.0110
	0.0211	—	**0.0083**	**0.0102**	**0.0097**	**0.0078**	**0.0080**	**0.0175**	**0.0149**
B	*0.0293*		*0.0092*	*0.0122*	*0.0125*	*0.0099*	*0.0090*	*0.0226*	*0.0178*
	0.0051		0.0077	0.0084	0.0048	0.0044	0.0062	0.0084	0.0103
	0.0209	**0.0203**	—	**0.0137**	**0.0133**	**0.0114**	**0.0063**	**0.0205**	**0.0189**
C	*0.0272*	*0.0280*		*0.0152*	*0.0156*	*0.0132*	*0.0069*	*0.0205*	*0.0216*
	0.0070	0.0033		0.0118	0.0081	0.0077	0.0051	0.0118	0.0136
	0.0304	**0.0282**	**0.0244**	—	**0.0119**	**0.0093**	**0.0128**	**0.0197**	**0.0184**
D	*0.0397*	*0.0387*	*0.0318*		*0.0137*	*0.0104*	*0.0142*	*0.0239*	*0.0210*
	0.0087	0.0049	0.0069		0.0088	0.0077	0.0103	0.0121	0.0140
	0.0291	**0.0277**	**0.0234**	**0.0313**	—	**0.0082**	**0.0125**	**0.0200**	**0.0177**
E	*0.0368*	*0.0364*	*0.0283*	*0.0394*		*0.0095*	*0.0149*	*0.0254*	*0.0210*
	0.0128	0.0084	0.0106	0.0121		0.0048	0.0066	0.0088	0.0107
	0.0194	**0.0194**	**0.0179**	**0.0268**	**0.0274**	—	**0.0108**	**0.0179**	**0.0151**
F	*0.0250*	*0.0270*	*0.0237*	*0.0361*	*0.0343*		*0.0127*	*0.0225*	*0.0173*
	0.0060	0.0023	0.0042	0.0059	0.0099		0.0062	0.0084	0.0103
	0.0255	**0.0237**	**0.0140**	**0.0285**	**0.0270**	**0.0219**	—	**0.0201**	**0.0184**
G	*0.0316*	*0.0315*	*0.0166*	*0.0352*	*0.0325*	*0.0277*		*0.0247*	*0.0213*
	0.0110	0.0069	0.0072	0.0105	0.0132	0.0083		0.0103	0.0121
	0.0542	**0.0526**	**0.0488**	**0.0543**	**0.0557**	**0.0517**	**0.0523**	—	**0.0109**
Gk[c]	*0.0778*	*0.0773*	*0.0694*	*0.0761*	*0.0766*	*0.0743*	*0.0722*		*0.0146*
	0.0118	0.0085	0.0105	0.0121	0.0158	0.0095	0.0144		0.0048
	0.0477	**0.0463**	**0.0425**	**0.0463**	**0.0474**	**0.0457**	**0.0466**	**0.0204**	—
Kd[c]	*0.0683*	*0.0685*	*0.0603*	*0.0655*	*0.0653*	*0.0665*	*0.0647*	*0.0254*	
	0.0140	0.0010	0.0120	0.0136	0.0179	0.0100	0.0160	0.0113	

[a] Overall divergences are Jukes-Cantor distances calculated using PAUP* 4.0; silent and replacement site divergences calculated using DnaSP 3.0 (Rozas and Rozas, 1999).

[b] For purposes of computing synonymous site divergences, the anonymous loci *A1341* and *G1134* were considered noncoding due to lack of significant BLASTX scores or large open reading frames. For nrITS, ITS-1 and ITS-2 regions were treated as synonymous sites, while 5.8S rDNA nucleotide positions were treated as nonsynonymous sites.

[c] Abbreviations for outgroup taxa are *Gk* = *Gossypioides kirkii* and *Kd* = *Kokia drynarioides*.

malense and *Gossypioides kirkii*, to a low of 0.0140 between the C- and G-genome representatives (Table 3). The average Jukes-Cantor distance among ingroup taxa was 0.0242 ± 0.0046, and the average distance between all ingroup taxa and the two outgroup taxa was 0.0494 ± 0.0041.

Combined chloroplast and nuclear sequences—Pairwise ILD test results from a comparison of the combined cpDNA vs. the nDNA data set revealed significant heterogeneity (P = 0.001). This heterogeneity is due primarily to the conflicting phylogenetic resolution of B- and E-genome representatives (discussed in ‘‘phylogenetic analysis’’ sections below). Because of the numerous possible biological sources for partition incongruence between chloroplast DNA and nuclear DNA evolution (Wendel and Doyle, 1998), we chose not to combine the cpDNA and nDNA data sets into a single data set for ‘‘total evidence’’ analysis.

Individual and combined indel data sets—Because of the large size of the combined cpDNA and nDNA data sets, a substantial number of indels were identified with high confidence of positional homology (Appendix 2; http://ajbsupp.botany.org/v89/Cronn/.doc). Using outgroup sequences to polarize changes in the ingroup, we scored 24 and 25 potentially phylogenetically informative indels in the cpDNA and nDNA data sets, respectively. The majority of indels scored in these data sets (38 of 49, or 77.6%) were contiguous gaps with identical 5′ and 3′ ends and hence likely represent the product of single indel events. These indels varied in length from a low of 1 bp to a high of 49 bp. The remaining 11 indels included six perfect duplications 4 bp or greater in length, and 5 indels of variable length with one conserved end (either the 5′ or 3′). In contrast to the combined cpDNA and nDNA nucleotide data sets, tests for incongruence between the two indel data sets using the ILD test did not return a significant value (P = 0.092). Accordingly, we chose to combine the cpDNA + nDNA indel data sets into a single indel data set composed of 49 characters.

Phylogenetic analysis of individual and combined data sets—*Phylogenetic inference from chloroplast genes*—Maximum likelihood topologies obtained for each of the individual chloroplast loci (*matK, ndhF, rpl16* intron and *trnT-trnL* spacer) are summarized in Fig. 3, and results for the combined cpDNA data set are shown in Fig. 4A. Analyses of the individual chloroplast loci yielded more than one MP tree in all cases. Since the single tree produced by ML was always included among the parsimony trees, the trees shown in Fig. 3 can be considered one representative example of the 5 trees (*ndhF*), 7 trees (*matK*), 32 trees (*rpl16* intron), and 6 trees (*trnT-trnL* spacer) recovered from MP analysis for individual chloroplast loci. Despite minor topological differences between ML and MP trees, Fig. 3 incorporates relevant information for both of these forms of inference, including branch lengths for ML trees, and descriptive statistics (tree lengths, consistency, and retention indices) and measures of support

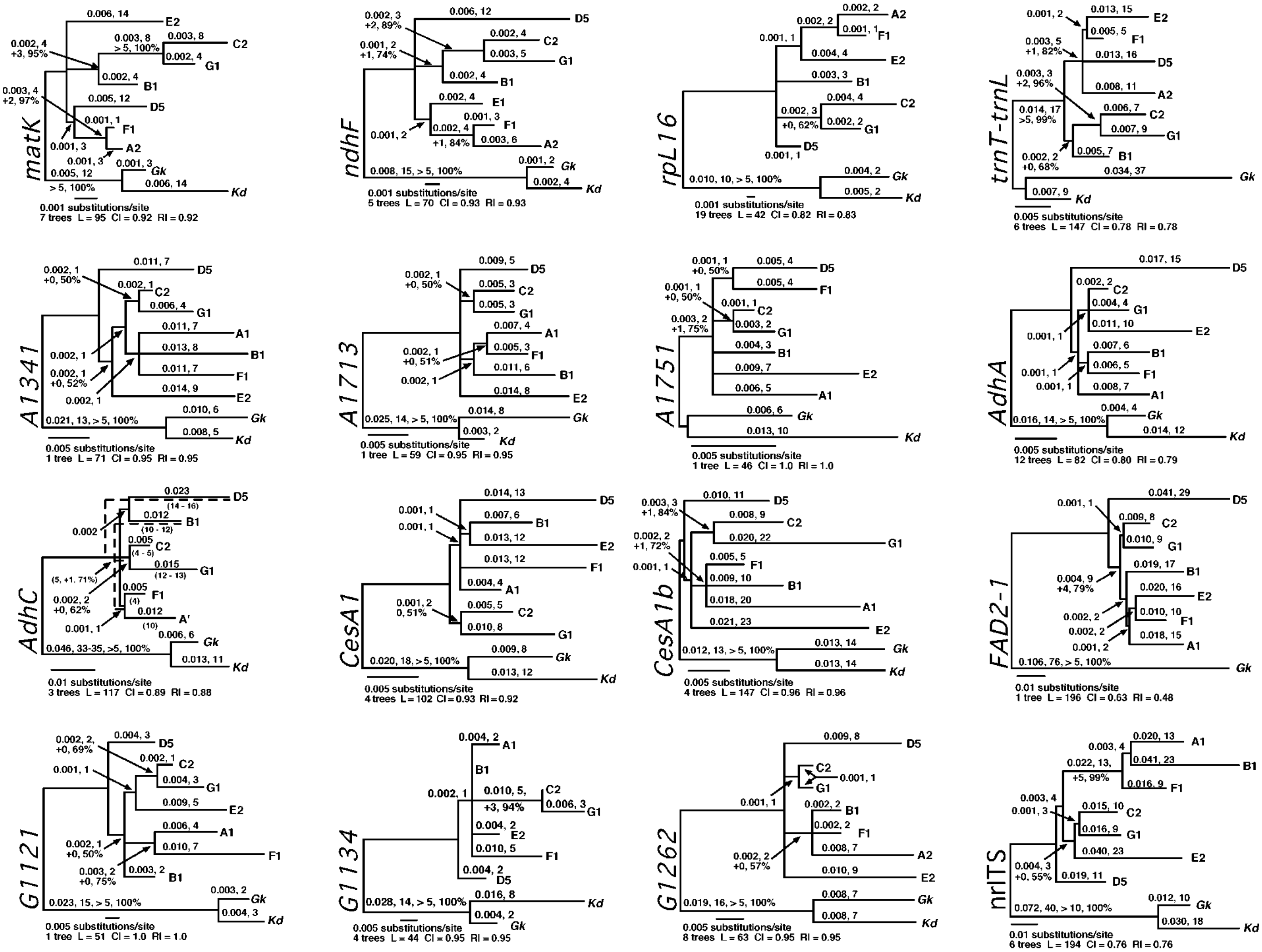

matK
0.001 substitutions/site
7 trees L = 95 CI = 0.92 RI = 0.92
ndhF
0.001 substitutions/site
5 trees L = 70 CI = 0.93 RI = 0.93
rpL16
0.001 substitutions/site
19 trees L = 42 CI = 0.82 RI = 0.83
trnT-trnL
0.005 substitutions/site
6 trees L = 147 CI = 0.78 RI = 0.78
A1341
0.005 substitutions/site
1 tree L = 71 CI = 0.95 RI = 0.95
A1713
0.005 substitutions/site
1 tree L = 59 CI = 0.95 RI = 0.95
A1751
0.005 substitutions/site
1 tree L = 46 CI = 1.0 RI = 1.0
AdhA
0.005 substitutions/site
12 trees L = 82 CI = 0.80 RI = 0.79
AdhC
0.01 substitutions/site
3 trees L = 117 CI = 0.89 RI = 0.88
CesA1
0.005 substitutions/site
4 trees L = 102 CI = 0.93 RI = 0.92
CesA1b
0.005 substitutions/site
4 trees L = 147 CI = 0.96 RI = 0.96
FAD2-1
0.01 substitutions/site
1 tree L = 196 CI = 0.63 RI = 0.48
G1121
0.005 substitutions/site
1 tree L = 51 CI = 1.0 RI = 1.0
G1134
0.005 substitutions/site
4 trees L = 44 CI = 0.95 RI = 0.95
G1262
0.005 substitutions/site
8 trees L = 63 CI = 0.95 RI = 0.95
nrITS
0.01 substitutions/site
6 trees L = 194 CI = 0.76 RI = 0.76

Fig. 3. Phylogenetic trees resulting from analyses of individual data sets. Taxa are designated by their genome designation. For each data set the ML (maximum likelihood) topology is shown. In all cases except one (*AdhC*), the ML topology was identical to one of the MP (maximum parsimony) trees recovered. The MP length (L), consistency index with autapomorphies excluded (CI), and the retention index (RI) are also shown. Branch lengths are indicated by a three decimal-place value representing the ML estimate, and an integer representing the MP-inferred branch length. Additional values on some branches indicate decay values (>0) and parsimony jackknife percentages (≥50%) for a given node. For *AdhC*, ML and MP analyses identified different optimal topologies. The ML analysis resulted in an ingroup topology as shown by solid lines connecting [(D + B)(C + G)(F + A′)]. The strict consensus of the three most parsimonious trees, as shown by dashed lines, had an ingroup topology of [D{(C + G)(B)(F + A′)}]. For this tree, the ML branch lengths are shown above the branches, while MP-inferred branch lengths are shown below the branches.

for MP trees (decay indices [*d*] greater than zero, and jackknife percentages [*j*] greater than 50%).

Results obtained from the *matK* and *trnT-trnL* spacer data sets generally agree with those previously obtained from cpDNA restriction site and *ndhF* gene sequence analyses (Fig. 1; Wendel and Albert, 1992; Seelanan, Schnabel, and Wendel, 1997). The six most-parsimonious trees obtained from *trnT-trnL* spacer show a basal dichotomy that divides *Gossypium* into two clades, one composed of the Australian C- and G-genomes plus the African B-genome and a second composed of African A-, E-, and F-genomes and the New World D-genome. This dichotomy is weakly supported, as the node supporting the B + (C + G) clade shows a jackknife support value of 68%. Similarly, decay analysis shows that this clade collapses in trees one step longer than the most parsimonious, yielding an ingroup trichotomy composed of a C + G clade, an A + E + F + D clade, and a solitary B-genome lineage. Strict consensus trees obtained from *matK* (seven trees) and *ndhF* (five trees) indicate an unresolved trichotomy in the genus, consisting of a B + (C + G) clade, an (A + F) clade, and a solitary D-genome clade (*ndhF*) or E-genome clade (*matK*). In contrast to *trnT-trnL* spacer data, the association between the B-genome and the (C + G)-genomes appears robust in *matK* and *ndhF*; the B + (C + G) clade survives three steps of decay in the *matK* data set (*j* = 95%), and one step of decay in the *ndhF* data set (*j* = 74%). In *matK* and *ndhF*, the African E-genome shows phylogenetic instability. The strict-consensus MP trees from *ndhF* place the E-genome in a clade sister to the A + F genomes (an association that survives one step of decay), while the strict consensus MP tree from *matK* fails to unite the E-genome with any of the African genome groups. Remarkably, only four potentially phyloge-

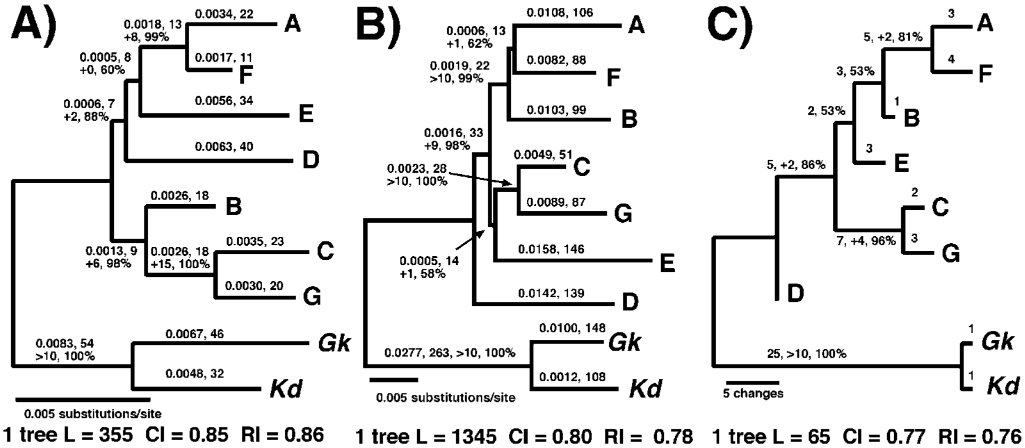


Fig. 4. Phylogenetic trees derived from combined data sets. (A) Maximum likelihood (ML) topology derived from the 7121-bp combined chloroplast data set; topology is identical to the single most-parsimonious tree recovered from cladistic analysis. (B) ML topology derived from the 10 850-bp combined nuclear data set; topology is identical to the single most-parsimonious tree recovered from cladistic analysis. (C) Single most-parsimonious tree derived from 49 phylogenetically-informative chloroplast and nuclear indels. Branch lengths are reported in substitutions per site [from ML in panels (A) and (B)], and as evolutionarily inferred steps from maximum parsimony in (A), (B), and (C). Decay values and jackknife support percentages are given for nodes where $d > 0$ and $j > 50\%$; Maximum parsimony (MP) length (L), consistency index with autapomorphies excluded (CI), and the retention index (RI) are also provided.

netically informative sites were recovered in the *rpl16* intron data set despite the fact that over 1100 bp were sequenced from each taxon; as expected, this gene provided little cladistic structure, as indicated by the 32 MP trees recovered.

In contrast with the inconsistent basal branching patterns, individual plastid sequences provide support for two sister-taxon pairs, an Australian clade composed of the (C + G)-genomes and an African clade composed of the (A + F)-genomes. All four cpDNA loci provide unanimous support for a (C + G)-genome clade, with decay and jacknife values ranging from robust in *matK* ($d = 5$, $j = 100\%$) to weak in the *rpl16* intron ($d = 0$, $j = 62\%$). Three of the four data sets also include the African B-genome in a monophyletic clade with the (C + G)-genomes, while the fourth data set (*rpl16* intron) simply lacks character support for this node. Three of the four loci analyzed (*matK*, *ndhF*, and *rpl16* intron) also showed support for a sister-taxon relationship between the A- and F-genome cottons (*ndhF*, $d = 1$ and $j = 84\%$; *matK*, $d = 1$ and $j = 97\%$; *rpl16* intron, $d = 0$ and $j < 50\%$). The *trnT-trnL* spacer data set fails to support this relationship, as the F-genome is weakly resolved ($d = 0$, $j < 50\%$) as sister to the E-genome. Aside from evidence supporting a monophyletic Australian (C + G)-genome clade (perhaps associated with the African B-genome lineage) and a sister relationship between the A- and F-genomes, little new phylogenetic insight emerged from the analysis of three individual cpDNA data sets. Nodes left unresolved by these analyses include the placement of the New World D-genome and African-Arabian E-genome, as these taxa exhibited affinities to either the (A + F)-genome clade (*matK* and *trnT-trnL* spacer for the D-genome; *ndhF*, *rpl16* intron and *trnT-trnL* spacer for the E-genome) or were left unresolved in a genus-wide polytomy (*ndhF* and *rpl16* intron for the D-genome; *matK* for the E-genome).

In sharp contrast to the four incompletely resolved individual chloroplast gene trees, ML and MP analyses of the concatenated cpDNA data sets (7121 included characters per taxon) yield a single most parsimonious tree that fully resolves the brancing order of the seven ingroup taxa (Fig. 4A). While the cpDNA tree combines features of individual cpDNA gene trees, it is important to note that no individual gene tree yielded a topology identical to that inferred from the combined cpDNA data set. This tree shows *Gossypium* to be divided into two primary lineages, with one clade composed of B + (C + G)-genomes, and the other composed of the D + [E + (A + F)]-genomes. The clade composed of the B + (C + G)-genome cottons is supported by nine synapomorphies and is strongly supported by decay analysis ($d = 6$) and jackknife resampling ($j = 98\%$); similarly, the cladistic affinity between the Australian (C + G)-genomes remains well supported ($d > 10$, $j = 100\%$). Conversely, there is only limited internal support for two of the three nodes included in the D + [E + (A + B)]-genome clade. Although the basal position for the D-genome ($d = 2$, $j = 88\%$) is relatively robust, the placement of the E-genome ($d = 0$, $j = 60\%$) is only weakly resolved. The sister relationship between the (A + F)-genomes is convincingly indicated, with high decay ($d = 8$) and jackknife ($j = 99\%$) support.

Phylogenetic inference from nuclear genes—Maximum likelihood topologies for individual nuclear genes are summarized in Fig. 3, and results for the concatenated nDNA data set are summarized in Fig. 4B. Superimposed on these ML trees are statistics and measures of support derived from parsimony analysis, as in nearly all cases the topology obtained from ML analysis did not conflict with the tree(s) obtained from MP analysis. The sole exception to this topological congruence was for *AdhC*, as ML and MP analyses identified different optimal topologies. In this case, the ML analysis resulted in an ingroup topology of [(D + B)(C + G)(F + A)], as indicated by bold lines in Fig. 3. The strict consensus of the three MP trees showed an ingroup topology of [D{(C + G)(B)(F + A)}] and is indicated by dashed lines. For this tree, ML branch lengths are shown above the branches, while MP-inferred branch lengths are shown in parentheses below the branches.

In a manner similar to cpDNA, 11 of 12 individual nuclear gene trees (all but *AdhA*) provide strong support for a sister-taxon association between the Australian (C + G)-cottons. The sister relationship indicated for the African (A + F)-genome cottons by cpDNA was less apparent in nuclear trees, as this clade appeared in only 3 of the 12 gene trees (*A1713, AdhC, G1121*); an equally common resolution, indicated by *A1341, CesA1b,* and *G1262,* was for the A- and F-genomes to be placed in an unresolved trichotomy that includes the B-genome. No locus positively contradicted the (A + F) grouping. Beyond the (C + G)-genome pair (and possibly an A + F-genome pair), trees obtained from individual (Fig. 3) and combined (Fig. 4B) nDNA data sets bear little resemblance to those obtained from cpDNA sequences (Figs. 3 and 4A). Features unique to the nuclear gene trees include: (1) placement of the D-genome as sister to the remainder of the genus, creating a basal ''New World/Old World'' split; (2) placement of the B-genome in an African clade consisting of (B + A + F)-genome cottons; and (3) a weak affinity of the African E-genome to the Australian (C + G)-genome cottons.

The most striking difference between trees based on cpDNA data (Wendel and Albert, 1992; Seelanan, Schnabel, and Wendel, 1997) and those obtained from nuclear sequences is that the latter repeatedly place the D-genome lineage sister to the remainder of the genus (Figs. 3 and 4B). This finding is supported by both ML and MP analyses of sequences from six different loci (*A1341, AdhA, FAD2-1, G1121, G1134,* and nrITS), and by parsimony analysis of *AdhC.* Four of the included nuclear loci (*A1713, A1751, AdhC,* and *G1262*) place the D-genome in an unresolved basal polytomy, a placement that does not conflict with a D-genome basal split. In fact, of the 12 nuclear loci sampled, only 1 (*CesA1*) shows a conflicting alternative for this earliest divergence in the genus. Results from *CesA1* provide weak support ($d = 0$) for a basal split between the Australian clade and the remainder of the genus, a result similar to that obtained from chloroplast *trnT-trnL* spacer data. While placement of the D-genome in a sister relationship to the remainder of the genus is only weakly supported among individual loci, the combined nDNA data set provides robust support for this placement (Fig. 4B). The combined Old World lineage shares a total of 33 inferred synapomorphies, and the node resolving New World D-genome cottons sister to Old World cottons has high support for decay ($d = 9$) and jackknife resampling ($j = 98\%$).

A second major difference between cpDNA-derived phylogenies and the trees obtained from nDNA concerns the placement of the African B-genome lineage (represented by *G. anomalum*). Chloroplast DNA sequence data resolve the B-genome as sister to the Australian (C + G)-genome cottons, rather than with other African cottons (Figs. 3 and 4A). In contrast, the nuclear data sets repeatedly unite the B-genome with one or both of the African A- and F-genome cottons. Five loci (*A1341, A1713, CesA1b, G1262,* and nrITS) independently support a monophyletic clade comprising the B-, A-, and F-genomes, while a sixth locus (*FAD2-1*) includes the Arabian E-genome in this assemblage. Although the (B + A + F)-genome clade fails to survive even a single step of decay in most data sets, the association is relatively robust in *CesA1b* ($d = 1$, $j = 72\%$) and nrITS ($d = 5$, $j = 99\%$). Three other loci reveal a close association between the B-genome and either the A-genome or F-genome, but do not support the inclusivity of a (B + A + F)-clade; the B-genome is sister to the A-genome without the F-genome in the *AdhA* data set, and it resolves as sister to the F-genome without the A-genome in the *AdhC* and *G1121* data sets. For two data sets, *AdhC* and *CesA1,* the B-genome resolves as sister to taxa other than A- or F-genome cottons, although these associations have no decay and little jackknife support. For *AdhC,* the B-genome is placed sister to the D-genome, but only in the ML analysis; this association is not seen in MP, as the B-genome falls into a polytomy composed of Old World cottons that are sister to the New World D-genome. Also, *CesA1* shows the B-genome to be the sister taxon of the E-genome, an association not evident in the other data sets. The preponderance of evidence (9 of 12 nuclear genes) indicates that the B-genome is closely allied to the A- and F-genome African cottons, and that these three taxa likely form a monophyletic lineage. Not surprisingly, analysis of the combined nDNA data set provided strong support for this association, as the preferred [B + (A + F)]-genome clade survives 12 steps of decay and yields jackknife resampling support of 99%.

As observed with cpDNA-derived phylogenies, trees obtained from nuclear genes show equivocal placement for the African-Arabian E-genome representative *G. somalense.* While plastid data generally show the E-genome to belong to a clade composed of African cottons (Fig. 4A), individual nuclear loci show that the E-genome is phylogenetically unstable, occupying alternative positions in the different gene trees (Figs. 3 and 4B). Three loci, *AdhA, G1121,* and nrITS, support an association of the E-genome with C- and G-genome cottons, albeit only weakly as only a single, nonhomoplasious character unites the E + (C + G)-genomes in each data set. Additionally, this node collapses after one step of decay and appears in only 58% of the jackknife replicates (Fig. 4B) in the combined nDNA data set. Alternative placements for the E-genome (all of which lack decay or jackknife support) include a basal position in the ''Old World'' cotton clade (*A1341*), a sister-taxon position with the B-genome (*CesA1),* as part of a monophyletic ''African clade'' that includes the A-, B-, and F-genome cottons (*FAD2-1* and nrITS) and as one of several lineages comprising a genus-wide, basal polytomy (*A1713, A1751, CesA1b, G1134,* and *G1262*). The position of the E-genome in the *AdhC* gene tree could not be determined, as amplification products could not be obtained from either *G. somalense* or *G. stocksii*. This locus has been partially eliminated (*G. arboreum*) or completely removed (*G. herbaceum*) from A-genome diploid cottons (Small et al., 1998; Small and Wendel, 2000), and it seems that E-genome *AdhC* may have experienced the same fate.

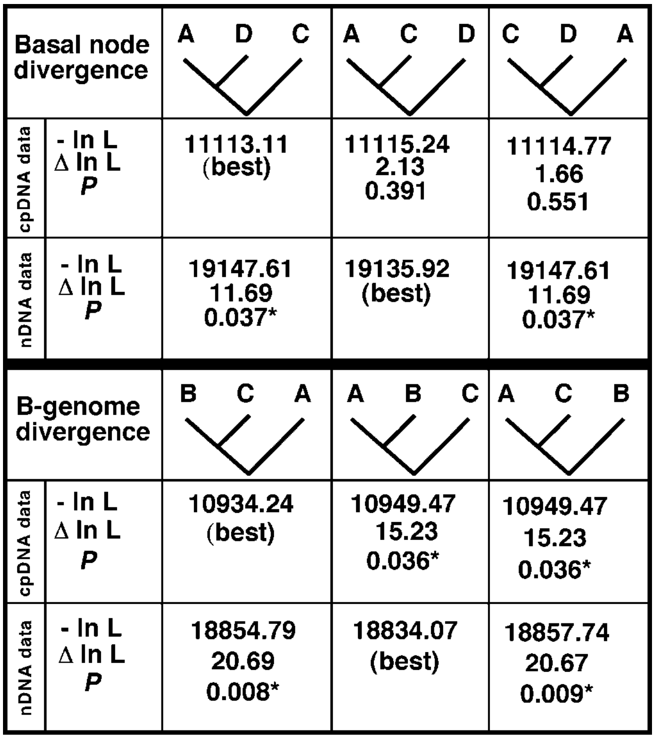
Phylogenetic inferences from indel events—Analysis of the combined cpDNA + nDNA indel data set produced a single most-parsimonious tree (length [L] = 65, retention index [RI] = 0.76; Fig. 4C) that is fully resolved. This tree shows greater overall similarity to the tree produced by the nDNA nucleotide data than by the cpDNA nucleotide data, as a New World/Old World cotton split represents the primary divergence event and because Old World taxa are divided into monophyletic Australian (C + G-genome) and African (A + B + E + F-genome) lineages. The primary divergence between the D-genome/Old World is supported by five synapomorphic indels (all from the nDNA data set), resulting in appreciable jackknife and decay support ($j = 86\%$, $d = 2$). A similarly robust node supports the Australian (C + G)-genome cotton clade ($j = 96\%$, $d = 4$), as four cpDNA and three nDNA indels mark this lineage. The final clade, composed entirely of African cottons, shows complete resolution but poor support values ($j = 53\%$, $d = 0$) at two of the three critical nodes.

While overall support for a monophyletic African clade (e.g., A-, F-, B-, and E-genomes) is low, inspection of cpDNA gap characters highlight an interesting inconsistency within the chloroplast data set. In the combined data set, five characters (*ndhF*-1, *trnT-trnL*-6, *trnT-trnL*-8 from cpDNA; *AdhA*-3 and *CesA1b*-3 from nDNA) provide support for placing the B-genome in the clade composed exclusively of African-Arabian taxa. In contrast, only a single character (*trnT-trnL*-7) from the cpDNA data set provides support for the combined B-genome–Australian clade. In this regard, indels reveal a remarkable feature of the cotton cpDNA data set, namely that chloroplast nucleotides and chloroplast indels inconsistently resolve the B-genome.

Statistical analysis of incongruence between nucleotide data sets—Despite the limited number of taxa included in this study, substantial topological disagreement was evident in comparisons between the cpDNA and nDNA nucleotide-derived trees (e.g., Fig. 4A vs. 4B). Indeed, agreement between the two topologies is limited only to sister relationships for the African A- and F-genomes and the Australian C- and G-genomes, associations that have been previously resolved using cpDNA (Wendel and Albert, 1992; Seelanan, Schnabel, and Wendel, 1997) and nrITS (Seelanan, Schnabel, and Wendel, 1997). Conflict among the remaining genome groups is most evident in the basal divergence event in the genus. The cpDNA tree reveals a basal divergence separating the common ancestor of the (C + G + B)-genomes from the common ancestor of the (D + E + F + A)-genomes. In contrast, nDNA indicates a geographically more coherent divergence, separating the ancestor of the New World D-genome from the ancestor of all Old World (A + F + B + E + C + G genomes) cottons. These two data sets show relatively robust character support for these alternative phylogenetic interpretations, as the basal node in the cpDNA tree (Fig. 4A) survives two steps of decay and exhibits a jackknife value of 88%, while the basal node from the nDNA tree (Fig. 4B) survives nine steps of decay and shows a jackknife value of 98%. Conflict is also evident in the affinity of the African B-genome, as represented here by *G. anomalum*. This taxon shows a strong affinity to African A- and F-cottons based on nDNA ($d > 10$, $j = 99\%$), yet it exhibits a closer relationship to the Australian (C + G)-genome lineage based on cpDNA ($d = 6$, $j = 98\%$).

To statistically evaluate the cpDNA and nDNA alternatives, we compared several topological variations using the Kishino-Hasegawa (KH) ΔlnL test. As might be anticipated from the strong character conflict between the cpDNA and nDNA data sets (indicated by the ILD test results), enforcing the alternative topology on either of these two concatenated data sets significantly increased the likelihood score, as indicated by the ΔlnL test: enforcing the nDNA-derived topology on the combined cpDNA data set resulted in a significant increase in the maximum likelihood score (unconstrained = 11 943.28; constrained to nDNA topology = 12 003.69; ΔlnL = 60.41; P = 0.0016), whereas even greater conflict appears when the nDNA data are forcibly constrained to the topology of the optimal cpDNA ML tree (unconstrained = 23 026.74; constrained to cpDNA topology = 23 099.55; ΔlnL = 72.81; P = 0.0001). These results indicate that the topological differences between the cpDNA and nDNA maximum likelihood trees are statistically significant and that the nDNA data set is more significantly perturbed by enforcing the alternative topology.

Additional topological tests were performed to evaluate whether alternative arrangements of selected taxa could be tolerated by the nDNA and cpDNA data sets. Because the testing of alternative, user-defined trees a posteriori may constitute a violation of the underlying assumptions of the KH test (Goldman, Anderson, and Rodrigo, 2000), we limited our comparisons to alternative topologies indicated from a priori phylogenetic hypotheses derived from cpDNA restriction site data (Wendel and Albert, 1992; Seelanan, Schnabel, and Wendel, 1997) and rDNA (Seelanan, Schnabel, and Wendel, 1997). To execute these tests, we pruned the number of taxa from nine (seven *Gossypium* and two outgroups) to a total of four (three ingroup and one outgroup) so that conflict among the most relevant lineages could be evaluated.



Basal node divergence		A D C	A C D	C D A
cpDNA data	- ln L Δ ln L *P*	11113.11 (best)	11115.24 2.13 0.391	11114.77 1.66 0.551
nDNA data	- ln L Δ ln L *P*	19147.61 11.69 0.037*	19135.92 (best)	19147.61 11.69 0.037*

B-genome divergence		B C A	A B C	A C B
cpDNA data	- ln L Δ ln L *P*	10934.24 (best)	10949.47 15.23 0.036*	10949.47 15.23 0.036*
nDNA data	- ln L Δ ln L *P*	18854.79 20.69 0.008*	18834.07 (best)	18857.74 20.67 0.009*

Fig. 5. Statistical evaluation of alternative *Gossypium* topologies using select taxa. Statistical support for each of three possible ingroup topologies based on the cpDNA or nDNA data sets was evaluated by computing Maximum likelihood (ML) scores (−ln L) for each user-input tree using the sequence parameters described in MATERIALS AND METHODS. The resulting increase in the ML score (Δln L) was evaluated for significance using the likelihood ratio test of Kishino and Hasegawa (1989). Comparisons that produced statistically significant scores ($P \leq 0.05$) are indicated by an asterisk. Upper panel: statistical support for alternative topologies involving the basalmost divergence event in the genus. Taxa selected to represent this divergence event include the A-, C-, and D-genome groups. Lower panel: statistical support for alternative topologies involving the phylogenetic affinity of the African B-genome lineage. Taxa selected to represent this divergence event include the A-, B-, and C-genome groups. The outgroup taxon for all analyses was *Gossypioides kirkii*.

To evaluate conflict concerning the primary divergence event among *Gossypium* genome lineages, we pruned the cpDNA and nDNA data sets to include the A-genome (*G. herbaceum*), C-genome (*G. robinsonii*), D-genome (*G. raimondii*), and one outgroup (*Gossypioides kirkii*). As three ingroup taxa are considered, there are three possible topologies for evaluation (Fig. 5, upper panel): (A + D) sister to C, which is the cpDNA topology reported by Wendel and Albert (1992), Seelanan, Schnabel, and Wendel (1997), and supported by our cpDNA analysis; (C + D) sister to A, the topology reported by Seelanan, Schnabel, and Wendel (1997) based upon nrITS using the distantly related taxon *Thespesia populnea* as an outgroup; and (A + C) sister to D, a topology that was indicated by analysis of the *Adh* multigene family in cotton (Small and Wendel, 2000) and is robustly supported by the nDNA data presented herein.

As shown in Fig. 5 (upper panel), the preferred topology for the cpDNA data set ([i.e., (A + D) sister to C]) yields an ML score that is statistically indistinguishable from the other two topological alternatives, i.e., (A + C) sister to D (P = 0.391) or (C + D) sister to A (P = 0.551). This result indicates that while ML and MP analyses resolve this node identically, character support for the preferred topology, a basal divergence of the Australian lineage from all other cottons, is limited. In contrast, the preferred topology for the nDNA data set [e.g., (A + C) sister to D] yields a maximum likelihood score that is statistically different from those obtained by the topological alternatives, i.e., (A + D) sister to C, or (C + D) sister to A (P = 0.037 in both cases). It should be noted that while we chose the A- and C-genomes to represent the African and Australian lineages in this analysis, identical results are obtained if the African F- and/or Australian G-genomes are substituted (results not shown).

Conflict in the placement of the B-genome lineage was evaluated in a similar manner by pruning the cpDNA and nDNA data sets to include the A-genome (*G. herbaceum*), B-genome (*G. anomalum*), C-genome (*G. robinsonii*), and one outgroup (*Gossypioides kirkii*). Hence, this analysis evaluates the strength of association between the B-genome and African vs. Australian cottons (Fig. 5, lower panel). With these included taxa, the three possible topologies are: (A + B) sister to C, the tree topology recovered from cpDNA by Wendel and Albert (1992), nrITS by Seelanan, Schnabel, and Wendel (1997), and from our analysis of nDNA; (B + C) sister to A, the tree topology recovered from cpDNA by Seelanan, Schnabel, and Wendel (1997), and supported by our analysis of cpDNA; and (A + C) sister to B, a topology that has yet to be proposed in literature and is unsupported by our analysis of nDNA and cpDNA.

As shown in Fig. 5, the preferred topology for the cpDNA data set, i.e., (B + C) sister to A, yields an ML score that statistically preferred to the other topological alternatives [e.g., (A + B) sister to C, or (A + C) sister to B; P = 0.036]. Similarly, the preferred topology for the nDNA data set [e.g., African (A + B) sister to Australian C] yields a ML score that is statistically more likely than those obtained from the topological alternatives, i.e., (B + C) sister to A, and (A + C) sister to B (P = 0.008 and 0.009, respectively). Again, while we chose the A- and C-genomes to represent the African and Australian lineages, identical results are obtained if the F- and/or G-genomes are used (results not shown).

The most revealing result from this exercise is the finding that placing the D-genome lineage sister to Old World cottons (in essence, the nDNA topology) results in an insignificant ΔlnL score in cpDNA data set (P = 0.391). This result suggests that while cpDNA and nDNA trees differ in their resolution of the basal divergence in *Gossypium*, placement of the D-genome is sufficiently equivocal in the cpDNA tree that a sister relationship between New World and Old World cottons is consistent with both data sets. Conversely, we find that the optimal placement of the B-genome, as indicated by ML analysis of cpDNA vs. nDNA, is statistically different in these data sets (Fig. 4A and 4B). Phylogenetic discord evident in the variable placement of the B-genome indicates that one of these data partitions has an evolutionary history that leads to phylogenetic incongruence. Events potentially involved include cytoplasmic introgression due to ancient hybridization, differential retention of ancestral chloroplast haplotypes, or poorly understood molecular-evolutionary phenomena that bias character evolution in a way that misleads phylogenetic reconstruction.

Timing the divergence of Gossypium lineages—To gain an understanding of the relative and absolute timing of the divergence events leading to the major extant cotton lineages, we used an ML approach to estimate internal and terminal branch lengths while enforcing a molecular clock constraint (Baum, Small, and Wendel, 1998; Sanderson, 1998). Maximum likelihood analyses of reduced data sets containing only synonymous sites showed that the cpDNA (4739 aligned positions) and nDNA (7978 aligned positions) synonymous site data sets deviated significantly from molecular clock expectations. Enforcing the clock constraint on the cpDNA data set yielded a significant increase in the ML score (−ln L unconstrained = 8402.92, −ln L clock enforced = 8438.71, Δln L = 35.79, $P = 1.9 \times 10^{-13}$); similarly, enforcing the clock constraint on the nDNA data set yielded a significant increase in the ML score (−ln L unconstrained = 17 926.08, −ln L clock enforced = 17 940.94, ΔlnL = 14.86, $P = 4.4 \times 10^{-5}$). By inspecting the trees visually and evaluating sequence heterogeneity using the 1D rate test of Tajima (Tajima, 1993), we found evidence of localized rate heterogeneity (results not shown). In particular, cpDNA from the C- and G-genomes showed elevated rates of divergence as compared to the B- and F-genomes. Similarly, nuclear sequences from the C-genome showed a significantly reduced rate of divergence as compared with nDNA from the A-, D-, and E-genome representatives (results not shown). By pruning the C-genome from the nDNA data set, reanalysis of the data using clock and nonclock models resulted in acceptance of the molecular clock (−lnL unconstrained = 17 500.87, −lnL clock enforced = 17 506.36; Δln L = 5.49; P = 0.089). Unfortunately, simple pruning of one taxon (B-, C-, F-, or G-genome, singly) or pairs of taxa (C + G, or B + F) from the cpDNA data did not produce data sets that met molecular clock expectations (data not shown). For this reason, only the nDNA data set, pruned of the C-genome, is used to estimate the timing of major divergence events between cotton genome groups.

To illustrate the temporal scale and rapidity of these divergence events, we estimated absolute divergence dates by applying molecular clock rates derived from nuclear genes of plant groups that are represented in the fossil record. Fossil-calibrated synonymous rates of *Adh* divergence for angiosperms vary about sixfold, with absolute rates ranging from 2.6×10^{-9} substitutions · synonymous site^{-1} · yr^{-1} for palms (Morton, Gaut and Clegg, 1996), 7.0×10^{-9} substitutions · synonymous site^{-1} · yr^{-1} for grasses (Gaut, 1998), and 1.5×10^{-8} substitutions · synonymous site^{-1} · yr^{-1} for *Brassica* (Koch, Haubold, and Mitchell-Olds, 2000). In the absence of *Gossypium* fossils, it is not possible to know which value is most appropriate in the present application. Nevertheless, this exercise provides an effective demonstration of the rapidity of cotton radiation, as discussed below, and permits us to consider the merits of earlier hypotheses regarding the timing of radiation of the major *Gossypium* clades.

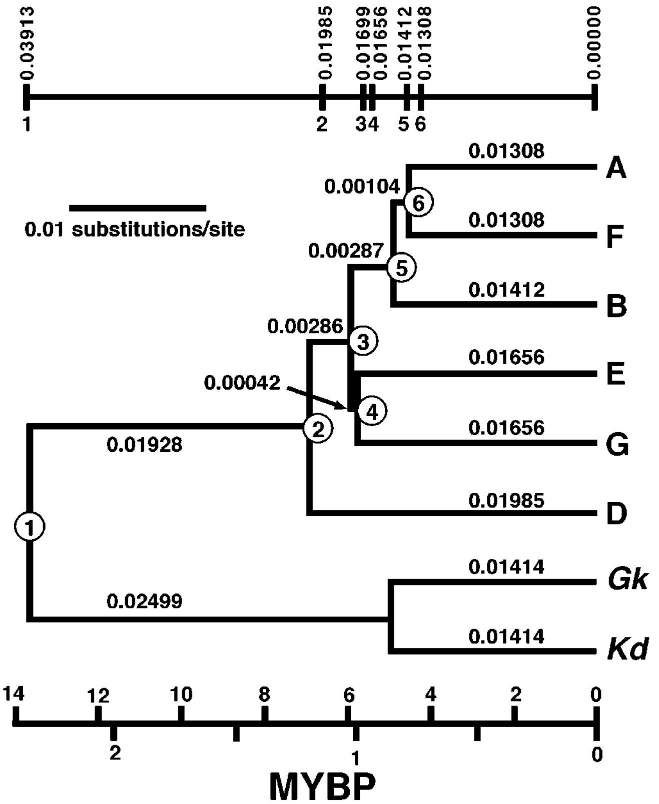


Fig. 6. Maximum likelihood (ML) tree obtained from nuclear synonymous site data (7978 bp) using a molecular clock constraint and inferences for absolute timing of divergence among diploid *Gossypium* lineages. Divergence events inferred by nodes 1–6 are superimposed on a relative scale of ML distances (upper bar), and an absolute timescale (in millions of years before present [MYBP]; lower bar) derived from synonymous site divergence estimates between diploid *Gossypium* and the outgroup taxa *Gossypiodes kirkii* (*Gk*) and *Kokia drynarioides* (*Kd*). Absolute divergence times are based on absolute rates of divergence estimated for palm *Adh* (upper time scale, 2.6 × 10^{-9} substitutions · synonymous site^{-1} · yr^{-1}; Morton, Gaut, and Clegg, 1996) and *Brassica Adh* (lower time scale, 1.5 × 10^{-8} substitutions · synonymous site^{-1} · yr^{-1}; Koch, Haubold, and Mitchell-Olds, 2000).

Average pairwise synonymous site divergence between the two outgroup species and all seven ingroup cottons was 0.0702 substitutions/synonymous site (standard deviation = 0.0055, N = 6521 synonymous sites per taxon). By using the lowest of the published *Adh* divergence rates (2.6 × 10^{-9} substitutions · synonymous site^{-1} · yr^{-1} for palms), we estimate that the separation between extant *Gossypium* diploids and *Gossypioides kirkii* (node 1, Fig. 6) occurred approximately 13.4 million years (MY) ago. This estimate provides a maximum likelihood-based conversion factor of 13.4 MY : 0.03913 ML units for the distance separating *Gossypium* from the combined outgroup *Gossypioides kirkii* + *Kokia drynarioides* (node 1 to terminals).

On the ML tree shown in Fig. 6, the basalmost radiation within *Gossypium* (i.e., divergence of the D-genome lineage from all cottons at node 2) is 0.01928 ML units distant from node 1 and 0.01985 ML units from terminal *Gossypium* taxa. The latter of these two distances (0.01985) approximates the relative age of the genus, although this estimate is conservative because extinct, *Gossypium*-like taxa may have diverged from extant cottons prior to this node. Using the conversion factor of 13.4 MY : 0.03913 ML units, the primary divergence event indicated by node 2 occurred approximately 6.8 MY ago (assuming a rate of 2.6 × 10^{-9} substitutions · synonymous site^{-1} · yr^{-1}). Lineages comprising the Old World clade (A-, B-, E-, F-, G-genomes) share an internal branch of 0.00286 ML units (nodes 2–3), a value that accounts for 14.4% of the age of the genus and corresponds to about 1.0 MY of common history. Members of the African clade (A-, B-, F-genomes) share an internal branch of 0.00287 ML units (node 3 to 5), a value that corresponds to 1.0 MY of common history. The shortest internal branches on the tree unite the A- and F-genomes (0.00104 ML units = 350 000 yr), and the E- and G-genomes (0.00042 ML units = 150 000 yr). If instead we use the faster of the published *Adh* rates (1.5 × 10^{-8} for *Brassica*), the separation between extant *Gossypium* taxa and *Gossypioides kirkii* (node 1) is reduced to approximately 2.3 MY ago. Using a conversion ratio of 2.3 MY : 0.03913 ML units, we can infer the separation of the D-genome lineage from the Old World lineage to have occurred about 1.2 MY ago (node 2). Members of the Old World clade exhibit 170 000 yr of shared history (nodes 2–3) and subsequently diverged to form their present-day genome lineages in the following 240 000 yr (nodes 3–6).

DISCUSSION

Rates of divergence and phylogenetic content of chloroplast and nuclear DNA in cotton—The pace and pattern of divergence in chloroplast and nuclear DNA sequences have been described for most diploid cotton species using chloroplast restriction site variation (Wendel and Albert, 1992), individual chloroplast gene sequences (Seelanan, Schnabel, and Wendel, 1997; Small et al., 1998), nuclear ribosomal DNA (Cronn et al., 1996; Seelanan, Schnabel, and Wendel, 1997), and low-copy nuclear sequences (Small et al., 1998; Cronn, Small, and Wendel, 1999; Seelanan et al., 1999; Small and Wendel, 2000; Liu et al., 2001). In general, these studies show that variation among species within genome groups is limited and that divergence between species from different genome groups far exceeds variation within genome groups. For example, pairwise cpDNA divergence among A-genome (two species) and D-genome (13 species) cottons is estimated at 0.00–0.32%, respectively (Wendel and Albert, 1992), while cpDNA divergence between A-genome and D-genome species averages 1.1%. Studies based on nuclear sequences reveal a similar trend, although nuclear divergences exhibit both a wider range of variation and higher overall levels of divergence than chloroplast DNA. For example, A- and D-genome cottons differ by as little as 0.74% (locus G1134; Cronn, Small, and Wendel, 1999) to as much as 23% (5S-rDNA; Cronn et al., 1996).

When averaged across all loci and ingroup taxa, our sample of twelve nuclear loci from diploid members of *Gossypium* evolve approximately 2.5-fold more rapidly than chloroplast sequences from the same species. When data sets are parsed into silent sites and nonsynonymous sites, silent rates among ingroup nuclear sequences are approximately 2.8-fold higher than silent sites in chloroplast sequences (mean K_S for nDNA = 0.0313; mean K_S cpDNA = 0.0113). Nonsynonymous divergence rates exhibit near equivalence between genome partitions, as ingroup nDNA replacement rates average only about 1.2-fold higher than cpDNA replacement rates (mean K_A for nDNA = 0.0080; mean K_A for cpDNA = 0.0069). These findings are consistent with prior studies in *Gossypium* (Small et al., 1998; Seelanan et al., 1999; Liu et al., 2001), and we note that these ratios of divergence between nuclear and chloroplast sequences from *Gossypium* are on the low end of the spectrum for angiosperms, which are typically reported as averaging around 5 : 1 (Gaut, 1998). The source of this low nuclear : chloroplast divergence ratio in *Gossypium* may reflect several factors, including (1) an unusually rapid pace of chloroplast divergence, (2) a slow rate of nuclear divergence, and/or (3) biases in gene sampling in our or other studies. We suspect that nuclear : chloroplast divergence ratios reflect factors that are both taxon-specific and sample dependent and that our observations from *Gossypium* are ''unusual'' only in that they include more genes than are commonly used in such comparisons.

One consequence of the 2.5-fold faster overall rate of divergence in the nuclear genome is that nuclear sequences could be expected to provide a corresponding increase in phylogenetic information compared with chloroplast sequences. This generalization is supported by analysis of five AD-genome tetraploid cotton species (Small et al., 1998), in that a single nuclear *AdhC* (1.6 kb) gene provided substantially greater phylogenetic resolution than four combined chloroplast loci exceeding 6 kb in length. Our results from the present study present somewhat of a paradox, as faster evolving individual nuclear genes do not improve phylogenetic resolution of cotton genome groups as compared with individual chloroplast genes of comparable length (Fig. 3). For example, fully resolved ML or MP topologies of the seven *Gossypium* diploid species should yield six dichotomously branching nodes. Individual ML analyses of the four chloroplast genes resolved an average of 4.75 nodes per tree, while 11 nuclear loci (excluding *AdhC,* which included one less taxon) resolved an average of 4.3 nodes per tree. While none of the four chloroplast gene trees were fully resolved by either MP or ML analysis, the nuclear genes yielded similar results, with only one locus (*FAD2-1* intron) yielding complete resolution of all nodes by both ML and MP analyses.

Our results indicate that nDNA, while evolving approximately 2.5-fold more rapidly than cpDNA in diploid cottons, may not yield a correspondingly higher proportion of phylogenetically informative characters (Table 2). While the proportion of variable sites within the ingroup is 2.5-fold higher in nuclear sequences (6.9%) than in chloroplast sequences (2.9%), the proportion of phylogenetically informative sites is only fractionally higher (0.83% vs. 0.70%). Restated, approximately one in four variable cpDNA sites (23.6%; Table 2) are potentially phylogenetically informative, while for nDNA this ratio is closer to one in nine (12.1%; Table 2).

Given the fact that the percentage of variable sites that are potentially phylogenetically informative (PPI) is over twice as high for cpDNA as nDNA, it is important to ask whether this reflects true synapomorphy or spurious homoplasy. Specifically, it is possible that the higher proportion of PPI sites in the cpDNA data is due to the presence of homoplasious sites that masquerade as synapomorphy. Such sites could artificially inflate the ratio of potentially phylogenetically informative to variable sites, leading to the erroneous conclusion that one data set was comparatively information rich. A high level of homoplasy may be unexpected for either of the included data sets, as ingroup sequences differ on average by only 1.03% (in cpDNA) to 2.42% (in nDNA); hence, one would not think that saturation phenomena would become an issue. Similarly, high homoplasy might be expected if base composition was sufficiently distorted such that substitutions were inherently biased due to high A + T or G + C content. In the present study, however, the sequences do not exhibit dramatically different levels of AT-richness; cpDNA contains 67% A + T and nDNA contains 61% A + T. Examination of nucleotide composition among the variable sites in these two data sets reveals a nearly identical base composition, with cpDNA showing 56.3% A + T (53.4% for parsimony informative sites) and nDNA showing 53.8% A + T (49.2% for parsimony informative sites). Hence, there is no obvious systemic reason for expecting more homoplasy in the cpDNA than nDNA data. It is possible, of course, that more obscure molecular evolutionary phenomena are operative and inaccessible to revelation through inspection. At present, however, we are unable to provide an explanation for the higher percentage of variable sites that are PPI in cpDNA than nDNA. We emphasize that notwithstanding this percentage difference, the absolute number of PPI sites is still nearly twice as large for nDNA ($N = 85$) than cpDNA ($N = 49$) data sets, due to the larger number of nucleotides sampled and the inherently 2.5-fold higher mutation rate.

A molecular phylogeny of Gossypium—Results from this study identify two alternative phylogenetic hypotheses for the *Gossypium* genome groups. The first hypothesis, revealed by approximately 7.0 kb of cpDNA (Fig. 4A), suggests that the first cladogenic event separated the common ancestor of the Australian (C + G)-genomes and African B-genomes from the common ancestor of the New World D-genome lineage and the African A-, E-, and F-genome lineages. In light of the hypothesized African origin for the genus *Gossypium* (Fryxell, 1979), this scenario indicates that the ancestor of extant cotton species initially divided into two major descendant African lineages. One of these major lineages further diverged in Africa, with one daughter lineage represented by modern-day B-genome cottons (distributed across central to southern Africa) and the second daughter lineage comprising the Australian C- and G-genomes (presently distributed across the Australian continent). The other major ancestral lineage split soon after formation, with one lineage originating following long-distance dispersal to the New World (the D-genome species) and the other clade containing the E-genome cottons (native to East Africa and the Arabian peninsula), F-genome cottons (native to East Africa), and A-genome cottons (native to East Africa and possibly the Indian subcontinent).

The second hypothesis, revealed by about 11 kb of nuclear DNA (Fig. 4B), differs from the cpDNA hypothesis with regard to the primary cladogenic event and cladistic relationships among African cottons. The nDNA phylogeny shows a primary divergence event that separates the New World D-genome lineage from the ancestor of all Old World taxa. Following this basalmost split in the genus, cottons comprising the Old World lineage divided into three clades, i.e., the Australian C- and G-genomes, the African-Arabian E-genome, and the African A-, B-, and F-genome cottons. Data are equivocal regarding branching order of these three clades, as alternative resolutions are weakly supported by the nDNA nucleotide (Fig. 4B) and overall indel (Fig. 4C) data. In contrast to the biogeographic scenario revealed by cpDNA, which diagnoses the origin of the genus as African, nDNA data fail to proffer an opinion regarding the ancestral home of the genus, as the basalmost divergence is between Old and New World clades.

While cpDNA- and nDNA-derived phylogenies show agreement with respect to sister relationships of some genome groups (A + F and C + G), they exhibit statistically significant incongruence regarding the primary divergence event in the genus and the phylogenetic affinities of the B- and E-genome African cottons. These incongruences may reflect either biological phenomena or peculiarities in the cpDNA and nDNA data sets. These various alternatives and possibilities are considered in the following section.

Primary divergence in the genus Gossypium—Perhaps the most obvious incongruence between the cpDNA and nDNA trees (Fig. 3) involves the primary divergence event in the genus. Parsimony and likelihood-based cpDNA topologies clearly resolve the combined Australian (C + G)-genomes and the African B-genome as sister to the remainder of the genus (Fig. 5). This resolution has moderate support ($d = 2$, $j = 88\%$) and corroborates earlier cpDNA-RFLP surveys (Fig. 1; Wendel and Albert, 1992). Perhaps of equal importance, the cpDNA topology shown in Fig. 4A is not contradicted by the constituent chloroplast loci, as ILD tests between the combined cpDNA data and individual gene data return no significant test statistics. Indeed, if nuclear genes had not been included in the present study, the absence of contradictory information within the plastome would have led us to conclude that the cpDNA-derived phylogeny (Fig. 4A) was internally consistent, fully resolved, and biologically correct.

This is not the case, however. Seven of twelve nuclear genes included in this study (*A1341, AdhA, CesA1b, FAD2-1, G1121, G1134,* and nrITS) directly contradict the cpDNA-based tree. Each of these independently specifies that the D-genome is sister to all Old World cottons (Fig. 4B). Thus, there is a striking incongruence between the cpDNA and nDNA genome partitions with respect to the initial divergence event in *Gossypium*. Resolution of this conflict requires a comparative evaluation of the characteristics and qualities of the two data partitions, a process that leads us to conclude that the nDNA topology most accurately reflects evolutionary history. Several reasons are provided for this conclusion.

First, multiple, independent data sets lead to the nDNA phylogeny. Among the nuclear loci sampled, 7 of 12 gene trees positively attest to the configuration captured in the global tree (Fig. 4B) vis-à-vis the earliest divergence, with only one locus (*CesA1*) weakly supporting an alternative. Although we sampled multiple regions from the chloroplast genome, these loci reside on a single nonrecombinant, haploid molecule. Accordingly, our sample of four chloroplast regions in essence represents four ‘‘domains’’ of a single genetic locus. In contrast, 10 of the 11 low-copy nuclear loci included in this study have been mapped to 7 of the 13 chromosomes of D-genome diploid cotton (*FAD2-1* has yet to be mapped, and nrITS may be present on as many as five chromosomes; Ji et al., 1999). Among the mapped loci, the closest locus pair is separated by a minimum of 25 centimorgans (cM) (*A1341* and *A1751*; Brubaker, Paterson, and Wendel, 1999). Accordingly, over evolutionary time, each nuclear locus provides an unlinked, independent estimate of *Gossypium* phylogeny. We find it particularly compelling that the majority of nuclear loci converge upon a single resolution (i.e., D-genome lineage is sister to all Old World lineages) and that there is a near-absence of conflict from other nuclear genes. To us, this constitutes strong evidence that the cpDNA data are positively misleading regarding the earliest radiation event in *Gossypium*.

A second reason for favoring the nDNA topology is that nucleotide character support for the basal-most divergence is greater in the nuclear tree than in the chloroplast tree. Inspection of parsimony-inferred branch lengths for the basalmost node reveals that 7 of 49 (14%) ingroup parsimony-informative sites in the cpDNA data are responsible for uniting the D-genome with the African A-, F-, and E-genomes. This value is considerably smaller than the corresponding value from nDNA, as parsimony analysis places 33 of the 85 (39%) ingroup parsimony-informative sites at this node. The net effect of this 4.7-fold difference in raw character support is that the nDNA topology is statistically preferred over other alternative topologies using the nDNA data set, whereas the cpDNA data fail to statistically discriminate among the topological alternatives (Fig. 5). The foregoing suggests that while the cpDNA and nDNA reconstructions differ in their resolution of the basalmost divergence, the cpDNA data are phylogenetically equivocal; hence, a sister relationship between New World and Old World cottons is consistent with both data sets.

A third and final reason for favoring the nDNA topology is that independent analysis of the indel data lends support to a basal position for the D-genome clade (Fig. 4C). Five separate indels support this basal split, and no indels positively contradict it (Appendix 2; http://ajbsupp.botany.org/v89/Cronn/.doc). Given this evidence as well as that discussed in the foregoing paragraphs, we conclude that the initial cladogenetic event in *Gossypium* separated Old and New World lineages and that the topological discord between cpDNA and nDNA in this respect simply reflects insufficient and/or misleading signal within the cpDNA data set.

An important implication of the foregoing is that the cpDNA data are positively misleading, despite the amount of sequence data generated and the higher percentage of variable sites that appeared potentially phylogenetically informative. The suggestion that emerges is that the cpDNA data may embed a sufficiently high level of homoplasy, relative to true synapomorphy, to allow evolutionary history to be erroneously inferred. In principle, insights into the nature of homoplasy may derive from inspection (Naylor and Brown, 1998), whereby changes in organellar DNA data (cpDNA, in this case) are mapped onto known phylogenies (the nDNA-inferred tree, in this case). In the present application, this exercise provides little insight into the errant behavior of cpDNA substitutions in *Gossypium*, as nucleotide positions supporting the basal resolution of the combined (B + C + G)-genome lineage are all noncoding and reside within either introns or spacers (results not shown).

Are African cottons monophyletic?—In the present study, the monophyly of the African-Arabian cottons is challenged by the cpDNA data in that the African, B-genome cottons appear cladistically sister to the (C + G)-genome Australian lineage (Figs. 3 and 4A). In addition, nDNA weakly suggest an affinity of the E-genome to the (C + G)-genome clade (Figs. 3 and 4B). In the most recent treatment of *Gossypium* (Fryxell, 1992), African cotton species from the A-, B-, E-, and F-genomes are united into the subgenus *Gossypium*. The putative monophyly of this taxonomic assemblage is not supported by obvious morphological synapomorphies. Indeed, it appears that all morphological character states used to define subgenus *Gossypium* (3–7 lobed, ovate to cordate leaves; embryos with prominent gossypol glands; corolla cream to yellow colored; epicalyx prominent, ovate-cordate and incised) are also variously present in species from the other two diploid subgenera (*Houzingenia*, the D-genome diploids, and *Sturtia*, comprising the C-, G-, and K-genome diploids). Perhaps the most compelling evidence for a monophyletic African clade comes from their present geographic distribution (native range includes Africa to Arabia), their shared, distinctively intermediate genome sizes (3.3–4.2 pg/2C vs. 2 pg/2C in subgenus Houzingenia and 5–7 pg/2C in subgenus Sturtia; see Wendel et al., 1999; Bennett, Bhandol, and Leitch, 2000), and chromosome pairing behavior in interspecific crosses (summarized in Endrizzi, Turcotte, and Kohel, 1985).

The chloroplast and nuclear DNA data presented here each provide a fully resolved picture regarding the relationships of African cottons relative to other lineages. There appears to be little question regarding the sister relationship between the African A- and F-genome cottons (Fig. 3), a relationship that was unsuspected (Fryxell, 1979, 1992) prior to the application of DNA data from chloroplast *ndhF* and nuclear nrITS sequences. Beyond this single, strongly supported point of agreement, however, the cpDNA and nDNA data sets fail to concur on the placement of the African B-genome and African-Arabian E-genome clades. Chloroplast DNA data place the B-genome lineage sister to the combined Australian (C + G)-genome lineage, a resolution supported by nine synapomorphic silent substitutions (two in *ndhF*, seven in spacer sequences) and three of the four chloroplast genes examined (all but *rpl16*, which was nearly devoid of phylogenetic signal). In addition, ML-based statistical tests of alternative resolutions for the B-, A-, and C-genomes (Fig. 5) show that the statistically preferred topology is one in which the B-genome is sister to the Australian lineage.

In striking contrast to the resolution indicated by cpDNA, the 10 988 bp isolated from 12 nuclear loci place the B-genome lineage solidly into an African clade that includes A- and F-genome cottons. This result is supported by 8 of 12 individual maximum likelihood gene trees (Fig. 3), and statistical tests of alternative resolutions for the B-, A-, and C-genomes show that the (B + A)-genome association is statistically preferred (Fig. 5). Previously published nrITS sequences (Seelanan, Schnabel, and Wendel, 1997) and low-copy gene sequences (locus *A1341*; R. Cronn, unpublished data) from the two remaining extant B-genome species, *G. capitis-viridis* and *G. triphyllum*, resolve all B-genome species as a monophyletic clade sister to the combined (A + F)-genome lineage. This resolution indicates that the anomalous resolution of our B-genome exemplar is a property of the entire B-genome lineage, not just the exemplar selected for inclusion in the present study. Based on the evolutionary independence of the estimates provided by 12 nuclear loci, we consider the nDNA resolution of the B-genome shown in Fig. 4B to be a more convincing reconstruction than that provided by the cpDNA data. This conclusion is additionally, albeit weakly, supported by the distribution of indels (Appendix 2; http://ajbsupp.botany.org/v89/Cronn/.doc), which suggest both African-Arabian monophyly and a cladistic relationship in which the B-genome is sister to the A + F clade (Fig. 4C). These results collectively suggest African-Arabian monophyly and by extension imply that geography, genome size, and cytogenetic pairing behavior are in the present circumstance more reliable predictors of phylogenetic relationships among African diploid cottons than are chloroplast gene sequences.

Among the various explanations for the incongruent cpDNA and nDNA resolutions of the B-genome are hybridization and homoplasy (Wendel and Doyle, 1998). The hybridization scenario entails reproductive contact between the progenitors of modern B-genome and Australian cottons, during which there was cytoplasmic introgression of an ‘‘Australian-like’’ plastome into the common ancestor of extant B-genome cottons. If this seemingly improbable event occurred, inspection of relative branch lengths in our combined cpDNA sequence data (Fig. 4A) indicates that it happened early in the history of the genus. Additionally, the cpDNA introgression would have had to occur prior to the diversification of modern-day B-genome species, as cpDNA restriction site analysis (Wendel and Albert, 1992) and limited *matK* sequencing (R. Cronn, unpublished data) of B-genome *G. capitis-viridis* and *G. triphyllum* show that all B-genome species share the same plastome haplotype. Finally, few bivalents are formed in B × C-genome interspecific hybrids (Phillips, 1966), just as in other combinations involving Australian cottons. Specifically, in B × C-genome hybrids, an average of 11 univalents are observed, indicating that modern B- and C- genome cottons are cytogenetically divergent, and are equally as divergent from the Australian C-genome cottons as are African A-genome (A × C = 10 univalents) and New World D-genome (C × D = 11 univalents) species.

Given the contradictory nuclear sequence data, the intercontinental allopatry of the relevant taxa, and cytogenetic evidence indicating poor meiotic pairing and complete sterility in intergenomic hybrids, it is difficult to accept cytoplasmic introgression as the preferred explanation for the phylogenetic incongruence regarding placement of the B-genome. Nonetheless, it remains a formal possibility. We note also that one B-genome species, *G. triphyllum*, native to Angola, Botswana, and Namibia, was previously placed in the Australian subgenus *Sturtia* (Fryxell, 1979) due to its relatively novel calyx form, pink-purple corolla pigmentation (similar to C-genome cottons), fruit pubescence (similar to the G-genome species *G. australe*), and leaf shape (trifoliolate, similar to the G-genome species *G. bickii*). These characteristics are unique among African cottons, and it is intriguing to consider that they may represent vestigial morphological remnants of an ancient hybridization between members of the Australian clade and the B-genome African lineage.

While ancient hybridization between progenitors of modern Australian and B-genome African clades may provide a tidy resolution for the incongruence between cpDNA and nDNA nucleotide data (and perhaps an explanation for the unique characteristics harbored within B-genome *G. triphyllum*), the actual source of this incongruence may be more obscure. Specifically, the conflicting resolutions may reflect the inability of cpDNA to accurately track evolutionary history in this recently diverged group. As noted above, this is suggested by the indel data, which we analyzed separately from the nucleotide data (Fig. 4C). In this analysis, most indels, even those from the chloroplast genome, support African-Arabian monophyly and the same cladistic resolution of the B-genome as indicated by the nDNA data set. Chloroplast indels supporting a monophyletic B-genome/African clade include (1) a two codon deletion from *ndhF* that unites B, A, and E; (2) a perfectly conserved 5 bp deletion in the *trnT-trnL* spacer that unites B and A; and (3) a 20-bp near-perfect tandem duplication in the *trnT-trnL* spacer that unites B, A, and E. In fact, of the 24 cpDNA indels scored in the chloroplast data set, only one (a conserved 7-bp insertion in *trnT-trnL*) unites the (B + C + G)-genome clade, even though six synapomorphic indels were found to unite the C- and G-genomes. In this respect, the cpDNA data are internally inconsistent, with nucleotides and indels indicating alternative resolutions of the B-genome. The underlying molecular evolutionary or perhaps nucleo-chemical basis of this discrepancy remain obscure, but this would seem to be an important avenue for future investigation given the ubiquitous application of cpDNA sequence data to systematic problems. As indicated previously, one rational approach for deducing the pattern of spurious homoplasy is to map changes in the cpDNA data onto ''known'' phylogenies (in the present case, the nDNA-inferred tree). Again, this exercise provides minimal insight into the errant placement of the B-genome lineage based on cpDNA, as the seven nucleotide positions supporting B + (C + G)-genome monophyly are either noncoding (five from introns or spacers) or occur at fourfold degenerate sites in exons (two from *ndhF*; results not shown).

Finally, we note a lack of agreement between cpDNA and nDNA regarding the placement of the E-genome lineage. In our opinion, this incongruence largely reflects a lack of character support caused by an inherently short interior branch. In our combined cpDNA data set, the E-genome is weakly ($d = 0$, $j = 60\%$) resolved as sister to the (A + F)-genome clade (Fig. 4A), whereas in the combined nDNA data set the E-genome is weakly suggested ($d = 0$, $j = 58\%$) to be sister to the C- and G-genome cottons (Fig. 4B). Collectively, the cpDNA and nDNA data sets essentially support an Old World trichotomy minimally composed of the (A + F)-genomes, the (C + G)-genomes, and a solitary E-genome. The apparent phylogenetic isolation of the E-genome exemplar *G. somalense* is likely shared among most if not all E-genome species, as parsimony analysis of cpDNA restriction sites (Wendel and Albert, 1992), nrITS sequences (Seelanan, Schnabel, and Wendel, 1997), and nuclear *A1341* sequences (R. Cronn, unpublished data) from the E-genome species *G. areysianum*, *G. somalense*, and *G. stocksii* reveal the E-genome group to be monophyletic. Additional supporting evidence for the relative genetic isolation of the E-genome lineage comes from meiotic chromosome pairing analysis in interspecific hybrids (Phillips, 1966). Intergenomic hybrids involving the E-genome show the greatest univalent frequency among all intergenomic hybrids studied, averaging 17 in E × A hybrids and 24 in E × C hybrids. The sequence data from 16 genes, therefore, do not strongly support African-Arabian monophyly per se, but instead highlight the rapidity with which the major Old World lineages diverged. We note that the E-genome resolves as sister to the B + A + F-genome clade in the independent analysis of indel data (Fig. 4C), lending weak support ($j = 53\%$) to African-Arabian monophyly. Finally, retention of the E-genome species in the subgenus *Gossypium* seems warranted on the basis of other features, such as geographic distribution, genome size, and the presence of abundant gossypol glands in embryos, a feature not shared by Australian C-, G-, and K-genome cottons (Fryxell, 1992).

Rapid cladogenesis and global dispersal of diploid Gossypium—Diploid cottons exhibit a nearly pantropical distribution and are native to every continent except Europe and Antarctica. The genus also exhibits remarkable morphological variation, some of which shows evidence of parallel gain (pubescence traits, corolla color, petal spot pigmentation, and leaf shape) in geographically and cytogenetically unrelated species (Fryxell, 1971, 1992). This combination of extensive morphological variation, morphological parallelism, and centers of diversity in Africa, Australia, and the New World led to hypotheses that (1) *Gossypium* was relatively ancient, with genome groups originating in the Cretaceous, 60–100 MY ago (Endrizzi, Turcotte, and Kohel, 1985; Endrizzi, Katterman, and Geever, 1989; Geever, Katterman, and Endrizzi, 1989) and (2) that the present distribution of New World and Old World lineages reflects divergence arising from the breakup of the Gondwanan supercontinent (see Saunders, 1961; Endrizzi, Turcotte, and Kohel, 1985). This hypothesis has been frequently cited even though it lacked support from alternative evidence such as macrofossils or palynological surveys. Indeed, the latter contradicts a Cretaceous diversification of the genus, as the oldest pollen referable to the Malvaceae is Eocene (38–46 MY ago) in age (Muller, 1981, 1984). At present, *Gossypium* fossils are limited to leaf prints in Hawaiian volcanic sediments dating to approximately 0.4 MY ago (Woodcock and Manchester, 1998), and fossil pollen has yet to be ascribed to the genus.

Molecular data, initially derived from nuclear DNA melting properties (c_ot curves; Endrizzi, Katterman, and Geever, 1989; Geever, Katterman, and Endrizzi, 1989), and more recently from chloroplast sequence variation (Wendel and Albert, 1992; Seelanan, Schnabel, and Wendel, 1997), have been used to evaluate divergence times among diploid cottons and to estimate the time of formation of the allopolyploid members of the genus. Using thermal stability measures of nucleotide divergence, Geever, Katterman, and Endrizzi (1989) provided the first estimate of divergence for members of *Gossypium*. Comparisons between *G. herbaceum* and *G. raimondii* indicated that the low copy DNA fraction differed by approximately 6% at the nucleotide level (see also Endrizzi, Katterman, and Geever, 1989). To estimate an absolute rate of divergence, these authors assumed that A- and D-genome lineages were separated approximately 100 MY ago, which was the latest date diploid *Gossypium* could have been continuously distributed across Africa, Australia, and South America. Their estimate of the absolute divergence rate, between 0.022–0.046%/MY, was ''markedly slower than that reported for other organisms'' (Geever, Katterman, and Endrizzi, 1989), as *Drosophila*, *Xenopus*, birds, primates, rodents, and ferns yield divergence rates considerably higher (0.17–0.4%/MY) using comparable methodologies.

Divergence dates have also been estimated using cpDNA using restriction site data (Wendel and Albert, 1992) and *ndhF* nucleotide sequences (Seelanan, Schnabel, and Wendel, 1997). Rather than assume a divergence time of 100 MY, these authors used absolute divergence rates measured from other taxa to calibrate a *Gossypium* cpDNA molecular clock and thereby evaluate proposals concerning divergence dates in the genus. Using cpDNA restriction site data, Wendel and Albert (1992) estimated the mean nucleotide divergence between C-genome cottons (representing one of two major cotton clades) vs. members of all other genome groups to be 1.66%. Using published *rbcL* rates, these authors calculated that the earliest radiation in *Gossypium* dated to between 24 and 33 MY ago. Direct sequencing of chloroplast *ndhF* (Seelanan, Schnabel, and Wendel, 1997) yielded lower overall estimates of nucleotide divergence among *Gossypium* genome groups, with the mean difference between C-genome cottons and other diploid genome groups measured at 0.86%. Using the same molecular rate calibration as Wendel and Albert, the primary divergence event in *Gossypium* was calculated at approximately 12 MY ago. Molecular clock-derived estimates, therefore, do not support the notion of a Cretaceous origin and diversification of the genus.

In this study, we applied synonymous site divergence data in nuclear DNA to evaluate the approximate timing of the origin of the cotton genus. This analysis supports the hypothesis of a relatively recent origin, as suggested by the earlier cpDNA data. Perhaps the present analysis is the most robust yet, as the inferences are based on the synonymous site divergences calculated from 7978 aligned sites. The mean divergence (K_S) between members of *Gossypium* is 0.0313, whereas the K_S between *Gossypium* and its sister genera *Gossypioides* and *Kokia* averages 0.0702, or approximately 2.3 times as high. When absolute divergence rates from palm and *Brassica* nuclear *Adh* are used to calibrate a molecular clock, the antecedant of modern-day *Gossypium* is estimated to have diverged from its relatives between 1.2 and 6.8 MY ago. Although these estimates cover a broad range, they are an order of magnitude lower than what might be predicted from the Cretaceous divergence predicted by Geever, Katterman, and Endrizzi (1989) and are more in line with previous cpDNA-based estimates. Indeed, the agreement between the present estimates and those of Seelanan, Schnabel, and Wendel (1997) is surprisingly high, particularly in light of the nature of the data sets employed and the different methods of clock calibration. Although there are many sources of error in the foregoing calculations, they offer useful approximations of both the absolute and relative dates of diversification events. Congruence among the estimates provides us with an added measure of confidence in suggesting that the genus has a relatively recent Miocene to Pliocene (2–13 MY ago) origin.

Evaluation of the absolute and relative timing of divergence events among *Gossypium* genome lineages highlights the rapidity to which this genus radiated into its present-day range. The clock-enforced nDNA ML tree shows that the divergence event separating *Gossypium* from the combined outgroup *Gossypioides kirkii* + *Kokia drynarioides* has a branch length of 0.03913 ML units (node 1 to terminals), a value that corresponds to between 13.4 MY ago (using palm *Adh* rates) to 2.3 MY ago (using Brassica *Adh* rates). The first cladogenic event among extant cottons, namely, separation of the D-genome lineage from the Old World (node 2), occurred 0.01985 ML units before present, a value that corresponds to 6.8–1.2 MY ago. All cotton genome groups (A–G, excluding C) then radiated within the next 0.00677 ML units, a value that corresponds to 34.1% of the age of the genus and 17.3% of the time since the separation of modern-day *Gossypium* and *Gossypioides/Kokia* lineages. Using absolute timescale estimates, this value (0.00677 ML units) indicates that all modern diploid lineages of cottons diverged within a time span of 2.3–0.4 MY.

In the absence of fossils it is impossible to determine which (if either) of the clock calibrations are closer to the truth, although we note that the life-history features of *Gossypium* species are more similar to those of members of the Araceae than the Brassicaceae. To the extent that generation time is correlated with molecular evolutionary rates, therefore, we suspect that the slower rates used above are more appropriate. We emphasize, however, two aspects of history that appear nearly incontrovertible based on the phylogenetic analysis and the evolutionary rate analyses: first, that the extant *Gossypium* lineages diversified sufficiently recently that they achieved their global distribution via an evolutionary history involving at least several long-distance, transoceanic dispersals and secondly, that the major cladogenetic events occurred on a temporally compressed scale relative to the age of the genus.

Summary comments on the promise of using multiple genes for resolving rapid radiations—Although this paper presents a novel phylogenetic hypothesis for the *Gossypium* genome groups, this hypothesis may be eclipsed in significance by the conclusion that multiple nuclear genes provide far greater resolution and insight than analyses based on cpDNA data alone (e.g., Wendel and Albert, 1992) or on cpDNA and rDNA analyses (e.g., Seelanan, Schnabel, and Wendel, 1997). We emphasize that if our present study had focused exclusively on sampling additional characters from either the chloroplast genome or the 45S ribosomal cistron, we would likely have gained fully resolved and competing estimates of the phylogeny of *Gossypium* genome groups, with no clear means of resolving the ensuing stalemate. By evaluating multiple, independent nuclear data sets, we gained insight into both the pattern and pace of divergence events leading to the evolution of the modern cotton lineages. This result is satisfying in that interrelationships among cotton genome groups have failed to be resolved for over 50 yr despite considerable scrutiny. In light of the rapidity and presumed recency of divergence events in *Gossypium*, perhaps it is not surprising that attempts using only a single line of evidence failed to resolve evolutionary relationships.

An additional, significant finding from this study is that cpDNA data lead to an incorrect picture of *Gossypium* evolutionary history. As we have accumulated more data, it has become apparent to us that cpDNA-nDNA incongruence is not limited solely to the analysis presented here, but is even more pronounced in an ongoing study of the New World, D-genome cotton species (R. Cronn et al., unpublished data). The underlying cause or causes of this incongruence remain unknown, but may include lineage sorting among polymorphic cpDNA haplotypes, frequent (and previously unsuspected) introgressive hybridization among now-allopatric species, or poorly understood molecular-evolutionary phenomena that bias character evolution in a way that confounds phylogenetic analysis (Wendel and Doyle, 1998). Irrespective of the source of the cpDNA-nDNA incongruence, the end result is that cpDNA frequently fails to accurately track the evolutionary history of diploid cottons. Because multiple nuclear loci are infrequently included in molecular-phylogenetic studies, it is unclear whether this is a problem unique to *Gossypium*, or, as seems more likely, whether this will turn out to be commonplace when numerous comparable studies are completed. The present study reveals a sobering and cautionary story, underscoring the truism that single gene-tree based (often cpDNA) phylogenetic hypotheses are subject to refutation and that these often will be refuted, at least in part, once additional data are gathered.

A comment is in order regarding the availability and use of ‘‘universal, low-copy’’ nuclear genes for phylogenetic studies. The present list of useful genes is growing at a rapid pace and will continue to expand rapidly in this era of comparative genomics. Our success in developing new markers for phylogenetic studies in *Gossypium* (a list that presently exceeds 60 genes; R. Cronn and J. Wendel, unpublished data) was rooted in a sizeable body of preexisting data on interrelationships of cotton species derived from morphology (Fryxell, 1971), cytogenetics (Phillips, 1966; Endrizzi, Turcotte, and Kohel, 1985), molecular linkage mapping (Brubaker, Paterson, and Wendel, 1999), and detailed gene family characterization provided by southern hybridization and expressed sequence tag (EST) database screening (e.g., Cronn, Small, and Wendel, 1999; Small, Ryburn, and Wendel, 1999). This wealth of information facilitates assessments of orthology and permits flexibility in choice of genes with desirable properties (e.g., gene length, chromosomal location, and gene structure). While similar information may exist for other plant groups (particularly crops and their wild relatives), the absence of these data for most taxa will require additional effort and experimentation to rigorously establish orthology for nuclear genes included in phylogenetic analyses.

In addition to this technical hurdle, we note that despite the wide rate heterogeneity of different nuclear genes (Wolfe, Li, and Sharp, 1987; Gaut and Doebley, 1997; Cronn, Small, and Wendel, 1999) and members of gene families (Small and Wendel, 2000), there is little assurance that a single nuclear marker will diverge at a rate appropriate for resolving phylogenetic issues. Additionally, functional redundancy may permit gene elimination in some members of a study, as is the case for *AdhC* in *Gossypium* (Small et al., 1998; Small and Wendel, 2000; this study). Finally, a history of temporally compressed divergence events, as suggested by the present study, will continue to vex attempts at confident phylogenetic resolution using any single gene. These cautionary comments highlight the challenges in developing multiple molecular markers for resolving what are often rapid phylogenetic radiations. The use of multiple, independent nuclear loci, however, promises not only to resolve phylogenetic relationships but offers a means by which hybridization events may be disentangled and insights may be gained into the rich and complex evolutionary history of flowering plants.

LITERATURE CITED

BALDWIN, B. G. 1997. Adaptive radiation of the Hawaiian silversword alliance: congruence and conflict of phylogenetic evidence from molecular and non-molecular investigations. *In* T. J. Givnish and K. J. Systma [eds.], Molecular evolution and adaptive radiation. Cambridge University Press, Cambridge, UK.

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

BENNETT, M. D., P. BHANDOL, AND I. J. LEITCH. 2000. Nuclear DNA amounts in angiosperms and their modern uses: 807 new estimates. *Annals of Botany* 86: 859–909.

BREMER, B., R. K. JANSEN, B. OXELMAN, M. BACKLUND, H. LANTZ, AND K.-J. KIM. 1999. More characters or more taxa for a robust phylogeny: case study from the coffee family (Rubiaceae). *Systematic Biology* 48: 413–435.

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

BRINKMANN, H., AND H. PHILIPPE. 1999. Archaea sister group of bacteria? Indications from tree reconstruction artifacts in ancient phylogenies. *Molecular Biology and Evolution* 16: 817–825.

BRUBAKER, C. L., A. H. PATERSON, AND J. F. WENDEL. 1999. Comparative genetic mapping of allotetraploid cotton and its diploid progenitors. *Genome* 42: 184–203.

BRUBAKER, C. L., AND J. F. WENDEL. 1994. Reevaluating the origin of domesticated cotton (*Gossypium hirsutum*; Malvaceae) using nuclear restriction fragment length polymorphisms (RFLPs). *American Journal of Botany* 81: 1309–1326.

BULL, J. J., J. P. HUELSENBECK, C. W. CUNNINGHAM, D. L. SWOFFORD, AND P. J. WADDELL. 1993. Partitioning and combining data in phylogenetic analysis. *Systematic Biology* 42: 384–397.

CRONN, R., R. L. SMALL, AND J. F. WENDEL. 1999. Duplicated genes evolve independently following polyploid formation in cotton. *Proceedings of the National Academy of Sciences, USA* 96: 14 406–14 411.

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

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

CUMMINGS, M. P., S. P. OTTO, AND J. WAKELEY. 1995. Sampling properties of DNA sequence data in phylogenetic analysis. *Molecular Biology and Evolution* 12: 814–822.

ENDRIZZI, J. E., F. R. H. KATTERMAN, AND R. F. GEEVER. 1989. DNA hybridization and the time of origin of three species of *Gossypium*. *Evolutionary Trends in Plants* 3: 115–119.

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

FARRIS, J. S. 1996. Parsimony jackknifing outperforms neighbor-joining. *Cladistics* 12: 99–124.

FARRIS, J. S., M. KALLERSJO, A. G. KLUGE, AND C. BULT. 1995. Testing significance of incongruence. *Cladistics* 10: 315–320.

FEHRER, J. 1996. Conflicting character distribution within different data sets on cardueline finches: artifact or history? *Molecular Biology and Evolution* 13: 7–20.

FRYXELL, P. A. 1971. Phenetic analysis and the phylogeny of the diploid species of *Gossypium* L. (Malvaceae). *Evolution* 25: 554–562.

FRYXELL, P. A. 1979. The natural history of the cotton tribe. Texas A & M University Press, College Station, Texas, USA.

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

GAUT, B. S. 1998. Molecular clocks and nucleotide substitution rates in higher plants. *In* M. K. Hecht [ed.] Evolutionary biology, vol. 30, 93–120. Plenum Press, New York, New York, USA.

GAUT, B. S., AND J. F. DOEBLEY. 1997. DNA sequence evidence for the segmental allotetraploid origin of maize. *Proceedings of the National Academy of Sciences, USA* 94: 6808–6814.

GEEVER, R. F., F. R. H. KATTERMAN, AND J. E. ENDRIZZI. 1989. DNA hybridization analyses of a *Gossypium* allotetraploid and two closely related diploid species. *Theoretical and Applied Genetics* 77: 553–559.

GOLDMAN, N., J. P. ANDERSON, AND A. G. RODRIGO. 2000. Likelihood-based tests of topologies in phylogenetics. *Systematic Biology* 49: 652–670.

HASEGAWA, M., H. KISHINO, AND T. YANO. 1985. Dating the human-ape split by a molecular clock of mitochondrial DNA. *Journal of Molecular Evolution* 22: 160–174.

HILLIS, D. M. 1996. Inferring complex phylogenies. *Nature* 383: 130–131.

HILLIS, D. M. 1998. Taxonomic sampling, phylogenetic accuracy, and investigator bias. *Systematic Biology* 47: 3–8.

HILLIS, D. M., J. P. HUELSENBECK, AND C. W. CUNNINGHAM. 1994. Application and accuracy of molecular phylogenies. *Science* 264: 671–677.

HUGHES, J. M., AND A. J. BAKER. 1999. Phylogenetic relationships of the enigmatic hoatzin (*Opisthocomus hoazin*) resolved using mitochondrial and nuclear gene sequences. *Molecular Biology and Evolution* 16: 1300–1307.

JI, Y., M. DEDONATO, C. F. CRANE, W. A. RASKA, M. NURUL ISLAM-FARIDI, T. D. MCKINGHT, H. J. PRICE, AND D. M. STELLY. 1999. New ribosomal RNA gene locations in *Gossypium hirsutum* mapped by meiotic FISH. *Chromosoma* 108: 200–207.

JOHNSON, L. A., AND D. E. SOLTIS. 1998. Assessing congruence: empirical examples from molecular data. *In* D. E. Soltis, P. S. Soltis, and J. J. Doyle [eds.], Molecular systematics of plants II: DNA sequencing, 297–348. Kluwer Academic Publishers, Dordrecht, The Netherlands.

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

KENNEDY, M., A. M. PATERSON, J. C. MORALES, S. PARSONS, A. P. WINNINGTON, AND H. G. SPENCER. 1999. The long and short of it: branch lengths and the problem of placing the New Zealand short-tailed bat, *Mystacina*. *Molecular Phylogenetics and Evolution* 13: 405–416.

KISHINO, H., AND M. HASEGAWA. 1989. Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data and the branching order in Hominoidea. *Journal of Molecular Evolution* 29: 170–179.

KOCH, M. A., B. HAUBOLD, AND T. MITCHELL-OLDS. 2000. Comparative evolutionary analysis of chalcone synthase and alochol dehydrogenase in *Arabidopsis*, *Arabis*, and related genera (Brassicaceae). *Molecular Biology and Evolution* 17: 1483–1498.

LARA, M. C., J. L. PATTON, AND M. N. F. DA SILVA. 1996. The simultaneous diversification of South American echimyid rodents (Hystricognathi) based on complete cytochrome b sequences. *Molecular Phylogenetics and Evolution* 5: 403–413.

LIU, Q., C. L. BRUBAKER, A. G. GREEN, D. R. MARSHALL, P. J. SHARP, AND S. P. SINGH. 2001. Evolution of the *FAD2-1* fatty acid desaturase 5′ UTR intron and the molecular systematics of *Gossypium* (Malvaceae). *American Journal of Botany* 88: 92–102.

MORTON, B. R., B. S. GAUT, AND M. T. CLEGG. 1996. Evolution of alcohol dehydrogenase genes in the palm and grass families. *Proceedings of the National Academy of Sciences USA* 93: 11 735–11 739.

MULLER, J. 1981. Fossil pollen records of extant angiosperms. *Botanical Review* 47: 1–142.

MULLER, J. 1984. Significance of fossil pollen for angiosperm history. *Annals of the Missouri Botanical Garden* 71: 419–443.

NAYLOR, G. J. P., AND W. M. BROWN. 1998. *Amphioxus* mitochondrial DNA, chordate phylogeny, and the limits of inference based on comparisons of sequences. *Systematic Biology* 47: 61–76.

NEI, M., S. KUMAR, AND K. TAKAHASHI. 1998. The optimization principle in phylogenetic analysis tends to give incorrect topologies when the number of nucleotides or amino acids used is small. *Proceedings of the National Academy of Sciences, USA* 95: 12 390–12 397.

OLMSTEAD, R. G., AND J. A. SWEERE. 1994. Combining data in phylogenetic systematics: an empirical approach using three molecular data sets in the Solanaceae. *Systematic Biology* 43: 467–481.

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

PERCIVAL, A. E., J. M. STEWART, AND J. F. WENDEL. 1999. Taxonomy and germplasm resources. *In* C. W. Smith and J. T. Cothren [eds.], Cotton; origin, history, technology and production, 33–63. John Wiley, New York, New York, USA.

PHILLIPS, L. L. 1966. The cytology and phylogenetics of the diploid species of Gossypium. *American Journal of Botany* 53: 328–335.

POE, S. 1998. The effect of taxonomic sampling on accuracy of phylogeny estimation: test case of a known phylogeny. *Molecular Biology and Evolution* 15: 1086–1090.

POE, S., AND D. L. SWOFFORD. 1999. Taxon sampling revisited. *Nature* 398: 299–300.

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

SANDERSON, M. J. 1998. Estimating rate and time in molecular phylogenies: beyond the molecular clock? *In* P. S. Soltis, D. E. Soltis, and J. J. Doyle [eds.], Molecular systematics of plants II: DNA sequencing, 242–264. Kluwer Academic Publishers, Dordrecht, The Netherlands.

SAUNDERS, J. H. 1961. The wild species of *Gossypium* and their evolutionary history. Oxford University Press, London, UK.

SEELANAN, T., C. L. BRUBAKER, J. M. STEWART, L. A. CRAVEN, AND J. F. WENDEL. 1999. Molecular systematics of Australian *Gossypium* section *Grandicalyx* (Malvaceae). *Systematic Botany* 24: 183–208.

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

SIMMONS, M. P., AND H. OCHOTERENA. 2000. Gaps as characters in sequence-based phylogenetic analysis. *Systematic Biology* 49: 369–381.

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

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

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

SOLTIS, D. E., M. E. MORT, M. W. CHASE, V. SAVOLAINEN, S. B. HOOT, AND C. M. MORTON. 1998. Inferring complex phylogenies using parsimony: an empirical approach using three large DNA data sets for angiosperms. *Systematic Biology* 47: 32–42.

STEELE, K. P., AND R. VILGALYS. 1994. Phylogenetic analysis of the Polemoniaceae using nucleotide sequences from the plastid gene *matK. Systematic Botany* 19: 126–142.

SWOFFORD, D. L. 2001. PAUP*: phylogenetic analysis using parsimony (*and other methods), version 4. Sinauer Associates, Sunderland, Massachusetts, USA.

TABERLET, P. L., L. GIELLY, G. PAUTOU, AND J. BOUVET. 1991. Universal primers for amplification of three non-coding regions of chloroplast DNA. *Plant Molecular Biology Reporter* 17: 1105–1109.

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

WAITS, L. P., J. SULLIVAN, S. J. O'BRIEN, AND R. H. WARD. 1999. Rapid radiation events in the family Ursidae indicated by likelihood phylogenetic estimation from multiple fragments of mtDNA. *Molecular Phylogenetics and Evolution* 13: 82–92.

WALSH, H. E., M. G. KIDD, T. MOUM, AND V. L. FRIESEN. 1999. Polytomies and the power of phylogenetic inference. *Evolution* 53: 932–937.

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

WENDEL, J. F., AND J. J. DOYLE. 1998. Phylogenetic incongruence: window into genome history and molecular evolution. *In* P. S. Soltis, D. E. Soltis, and J. J. Doyle [eds.], Molecular systematics of plants II: DNA sequencing, 265–296. Kluwer Academic Publishers, Dordrecht, The Netherlands.

WENDEL, J. F., A. SCHNABEL, AND T. SEELANAN. 1995a. Bidirectional interlocus concerted evolution following allopolyploid speciation in cotton (*Gossypium*). *Proceedings of the National Academy of Sciences, USA* 92: 280–284.

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

WENDEL, J. F., R. L. SMALL, R. C. CRONN, AND C. L. BRUBAKER. 1999. Genes, jeans, and genomes: reconstructing the history of cotton. *In* L. W. D. van Raamsdonk and J. C. M. den Nijs [eds.], Plant evolution in man-made habitats, 133–159. Proceedings of the VIIth International Organization of Plant Biosystematists. Hugo de Vries Laboratory, Amsterdam, The Netherlands.

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

WOODCOCK, D. W., AND C. A. MANCHESTER. 1998. Fossil cotton from the Salt Lake Crater area, O'ahu, Hawai'i. Records of the Hawaii Biological Survey for 1997, part 2. *In* N. L. Evenhuis and S. E. Miller [eds.], Bishop Museum occasional papers, 17–19. Bishop Museum Press, Honolulu, Hawaii, USA.