Rate Variation Among Nuclear Genes and the Age of Polyploidy in *Gossypium*

David S. Senchina, Ines Alvarez,* Richard C. Cronn,† Bao Liu,‡ Junkang Rong,§ Richard D. Noyes,|| Andrew H. Paterson,§ Rod A. Wing,¶ Thea A. Wilkins,# and Jonathan F. Wendel**

*Department of Botany, Iowa State University; †Pacific Northwest Research Station, USDA Forest Service; ‡Institute of Genetics and Cytology, Northeast Normal University, Changchun, China; §Plant Genome Mapping Laboratory, University of Georgia, Riverbend Research Center; ||Department of EPO Biology, University of Colorado; ¶Department of Plant Sciences, University of Arizona; and #Department of Agronomy and Range Science, University of California-Davis

Molecular evolutionary rate variation in *Gossypium* (cotton) was characterized using sequence data for 48 nuclear genes from both genomes of allotetraploid cotton, models of its diploid progenitors, and an outgroup. Substitution rates varied widely among the 48 genes, with silent and replacement substitution levels varying from 0.018 to 0.162 and from 0.000 to 0.073, respectively, in comparisons between orthologous *Gossypium* and outgroup sequences. However, about 90% of the genes had silent substitution rates spanning a more narrow threefold range. Because there was no evidence of rate heterogeneity among lineages for any gene and because rates were highly correlated in independent tests, evolutionary rate is inferred to be a property of each gene or its genetic milieu rather than the clade to which it belongs. Evidence from approximately 200,000 nucleotides (40,000 per genome) suggests that polyploidy in *Gossypium* led to a modest enhancement in rates of nucleotide substitution. Phylogenetic analysis for each gene yielded the topology expected from organismal history, indicating an absence of gene conversion or recombination among homoeologs subsequent to allopolyploid formation. Using the mean synonymous substitution rate calculated across the 48 genes, allopolyploid cotton is estimated to have formed circa 1.5 million years ago (MYA), after divergence of the diploid progenitors about 6.7 MYA.

Introduction

Molecular sequence data are widely employed as molecular clocks to address questions of the absolute and relative ages of various divergence events. A number of methodological and biological phenomena may affect the accuracy of molecular clock–based estimates (Gaut 1998; Hillis, Moritz, and Mable 1996; Sanderson 1998; Soltis et al. 2002). Major issues include rate heterogeneity caused by various evolutionary factors (e.g., generation time, germ-line processes, metabolic rate); difficulties in interpreting fossil or biogeographic data against which a clock might be calibrated; use of nonindependent lineages for calibration; and the inappropriate usage of calibrations on lineages for which they were not intended. One of the most significant factors is rate variation among genes, even at putatively neutral sites. Early surveys based on a dozen or fewer genes reported only a twofold to threefold variation in synonymous substitution rates (Wolfe, Sharp, and Li 1989; Gaut 1998), but recent studies reveal higher levels of intergenic rate variation (Kusumi et al. 2002; Tiffin and Hahn, 2002; Zhang, Vision, and Gaut 2002). Because rates of nuclear gene evolution within an organism vary so widely, sampling only one or a few genes in molecular clock applications can lead to high variance in divergence time estimates when only a single clock calibration is employed. A reasonable strategy to minimize this problem would be to sample multiple genes so that an average rate may be estimated. This is not to claim that other important sources of error do not exist; in particular, clock calibration and lineage-specific effects remain vexing and, to a certain extent, intangible problems. However, minimizing the issue of intergenic rate variation would appear to be an achievable and worthwhile goal. In the present study, we applied this rationale to a model system from *Gossypium* (cotton), employing sequence variation at 48 nuclear genes.

Gossypium L. contains 50 species whose phylogenetic relationships have been explored using multiple molecular data sets (Seelanan, Schnabel, and Wendel 1997; Small et al. 1998; Cronn et al. 2002*b*). Data indicate that shortly after its origin, *Gossypium* experienced rapid divergence (Cronn et al. 2002*b*), leading to modern monophyletic lineages that vary in chromosome size and interfertility (so-called "genome groups" A through G and K). There are five natural polyploids in the genus, which apparently spawned from a single polyploidization event 1 to 2 MYA (Cronn et al. 1996; Small et al. 1998; Wendel and Cronn 2003). All are "AD" genome tetraploids, combining an A-genome donated by the maternal diploid parent at the time of polyploid formation and a D-genome from the pollen parent (Galau and Wilkins 1989; Wendel 1989; Wendel and Cronn 2003). Among extant species, *G. herbaceum* L. and *G. arboreum* L. are the closest relatives of the A-genome progenitor, and *G. raimondii* Ulbrich is the best model of the D-genome progenitor (reviewed in Wendel and Cronn 2003). These two genome groups diverged from each other early in the evolution of the genus, perhaps 7 to 11 MYA (Seelanan, Schnabel, and Wendel 1997; Cronn et al. 2002*b*). Notably, the A-genome is about twice the size of the D-genome (2C = 3.8 pg versus 2.0 pg), and these size differences are perpetuated in the natural polyploids, which exhibit an additive genome size (Wendel et al. 2002 and references therein). Economically important polyploid species include *G. barbadense* L. (Sea Island and Pima cotton) and *G. hirsutum* L. (upland cotton).

Because of the absence of a fossil record, divergence times in *Gossypium* have been estimated primarily using molecular clock assumptions. These estimates have been based either on thermal denaturation-renaturation studies

Key words: *Gossypium*, cotton, polyploidy, molecular clock, substitution rates, evolution.

E-mail: jfw@iastate.edu.

Mol. Biol. Evol. 20(4):633–643. 2003
DOI: 10.1093/molbev/msg065

Table 1
Evolutionary Properties of Annotated Cotton Gene Sequences

Name	Gene Function	GenBank Accession Numbers	Sequence Properties			*K* Values[a]		
			Aligned Length	Replacements	Silent Sites	K_a	K_{sil}	K
A6 (Ea0003M07f)	Unknown	AF517646–AF517650	522	0	448	0.000	0.052	0.046
B5 (Ea0008D22)	Germin E protein precursor	AY116167–AY116171	669	7	226	0.014	0.061	0.030
B7 (Ea0010C09)	Polyubiquitin	AY117055–AY117059	616	0	138	0.000	0.068	0.015
B8 (Ea0010N12)	CAAX-prenyl protease	AY115496–AY115500	489	2	363	0.004	0.055	0.042
C3 (Ea0014A21)	Syntaxin	AY117060–AY117064	434	4	269	0.014	0.038	0.029
C7 (Ea0015F06)	SAH7 protein	AY117065–AY117069	839	28	464	0.073	0.065	0.068
D1 (Ea0017C01)	LIM-domain transcription factor	AY117110–AY117114	355	2	255	0.008	0.048	0.056
D2 (Ea0017H13)	IAA-responsive protein 9	AY116162–AY116166	522	9	233	0.017	0.036	0.025
D5 (Ea0017N07)	Transporter protein	AY117070–AY117074	909	20	269	0.027	0.046	0.032
D7 (Ea0018G05)	Root hair defective 3 (RHD3)	AY117080–AY117084	864	6	647	0.035	0.049	0.046
E1 (Ea0021I11)	Flavonoid 3′,5′-hydroxylase	AY117085–AY117089	361	7	142	0.020	0.066	0.038
E3 (Ea0021L07)	RGA1 protein	AY117075–AY117079	372	3	88	0.005	0.051	0.016
E5 (Ea0022I11)	Pollen surface protein	AY117090–AY117094	772	15	254	0.023	0.057	0.035
E9 (Ea0023A19)[b]	Kinase-associated protein	AF521968–AF521972	394	3	101	0.007	0.079	0.035
E11 (Ea0023B18)	Beta-D-glucan exohydrolase–like protein	AF517656–AF517660	223	5	109	0.027	0.057	0.042
F4 (Ea0024M11)	Ethylene receptor	AY117095–AY117099	366	6	156	0.012	0.038	0.023
F8 (Ea0025J02)	Sugar transporter	AY117100–AY117104	359	2	254	0.013	0.054	0.043
F12 (Ea0025P13)[c]	COP-1 interacting protein 7	AY116157–AY116161	287	13	94	0.052	0.078	0.060
G3 (Ea0026N16)	Quinone oxidoreductase	AY117105–AY117109	305	3	182	0.017	0.034	0.028
G8 (Ea0028I12)[b]	Flavonoid 3′-hydroxylase	AF525941–AF525945	144	3	32	0.023	0.075	0.034
G11 (Ea0029C08)	Sulfate transporter	AY116152–AY116156	344	3	143	0.004	0.085	0.038
H12 (Ea0033L21)	IAA-Ala hydrolase	AF517651–AF517655	288	10	125	0.040	0.071	0.054
A1286[de]	Unknown	AF136808–AF136811, AF201876	294	n/a	n/a	n/a	0.053	n/a
A1341[de]	Unknown	AF136813–AF136816, AF201877	666	n/a	n/a	n/a	0.042	n/a
A1520[d]	Unknown	AF136818–AF136821, AF201878	957	n/a	n/a	n/a	0.073	n/a
A1550[e]	Aldehyde dehydrogenase	AF201889–AF201893	1446	4	1111	0.011	0.058	0.052
A1623	Unknown	AF139474–AF139477, AF201879	719	14	425	0.041	0.039	0.040
A1625[d,e]	Unknown	AF139417–AF139470, AF201880	1048	n/a	n/a	n/a	0.035	n/a
A1713[d,e]	Unknown	AF139442–AF139445, AF201881	593	n/a	n/a	n/a	0.050	n/a
A1751	Subtilisin-like protease	AF139437–AF139440, AF201883	807	8	219	0.009	0.047	0.019
A1834[e,f]	Putative alpha-mannosidase	AF139452–AF139459	882	n/a	n/a	n/a	n/a	n/a
G1121[e]	Unknown	AF139432–AF139435, AF201884	749	3	439	0.006	0.050	0.032
G1134[d,e]	Unknown	AF139427–AF139430, AF201882	546	n/a	n/a	n/a	0.040	n/a
G1262[e]	P-glycoprotein	AF061085–AF061087, AF061089, AF201885	888	14	883	0.013	0.117	0.037
AdhA	Alcohol dehydrogenase A	AF085064, AF090146, AF136458–AF136459	951	3	467	0.007	0.055	0.031
AdhB[e]	Alcohol dehydrogenase B	AF226632–AF226635	1534	17	187	0.020	0.162	0.028
AdhC[b]	Alcohol dehydrogenase C	AF036568–AF036569, AF036574–AF036575	1680	15	868	0.018	0.094	0.053
AdhD[e]	Alcohol dehydrogenase D	AF059418, AF250203–AF250205	1556	14	1042	0.023	0.059	0.048
AdhE[e,f]	Alcohol dehydrogenase E	AF250208–AF250211	1629	n/a	n/a	n/a	n/a	n/a
CesA1	Cellulose synthase	AF139442–AF139445, AF201886	4025	20	1808	0.007	0.049	0.026
CesA2	Cellulose synthase	AF139447–AF139450, AF201887	2235	12	857	0.004	0.087	0.037
CLK1	Protein-associated kinase	AY124072–AY124076	2293	10	1580	0.016	0.048	0.041
MYB1	R2R3-MYB transcription factor 1	AY115501–AY115505	1066	12	342	0.013	0.101	0.040
MYB2	R2R3-MYB transcription factor 2	AY115506–AY115510	551	11	174	0.026	0.045	0.032
MYB3	R2R3-MYB transcription factor 3	AF377307–AF377308, AF377316, AF377318, AY115511	662	12	192	0.021	0.081	0.040
MYB4	R2R3-MYB transcription factor 4	AY115522–AY115524 (A1, D5, Dt)	385	11	120	0.025	0.097	0.050
MYB5	R2R3-MYB transcription factor 5	AY115512–AY115516	1057	16	366	0.019	0.052	0.040

Table 1
Continued

Name	Gene Function	GenBank Accession Numbers	Sequence Properties			*K* Values[a]		
			Aligned Length	Replacements	Silent Sites	K_a	K_{sil}	K
MYB6	R2R3-MYB transcription factor 6	AYY115517–AY115521	1032	17	182	0.018	0.018	0.018
Totals			40685	364	16254	n/a	n/a	n/a
Arithmetic mean			848	9	406	0.018	0.061	0.037
Median			639	9	254	0.017	0.055	0.038
Standard deviation			678	6.5	404	0.014	0.025	0.018

[a] K_a, K_{sil}, and K denote substitutions at nonsynonymous sites, synonymous sites within codons plus all noncoding positions, and rate of substitution across all sites, respectively. Numbers given are means for *Gossypium* versus outgroup comparisons.
[b] One or more taxa contains a pseudogene for the locus indicated.
[c] *Kokia kauiensis* was used as an outgroup instead of *Gossypioides kirkii*.
[d] Sequence treated as noncoding due to weak evidence for presence of exons.
[e] Not detected in cotton fiber cDNA library and hence classified as "nonfiber."
[f] No outgroup sequence available.

or on a small number of genes (Endrizzi, Katterman, and Geever 1989; Wendel 1989; Seelanan, Schnabel, and Wendel 1997; Cronn, Small, and Wendel 1999; but see Cronn et al. 2002*b*). Our objective here was to examine the extent of nuclear gene rate variation among genes and lineages in *Gossypium* and to use the resulting data set to generate a clearer understanding of the temporal components of the evolutionary history of *Gossypium*. We also wished to explore patterns of gene evolution in polyploid cotton, using as a comparative framework orthologs from the diploid progenitors. Previous studies (Cronn, Small, and Wendel 1999; Small and Wendel 2000) suggested that rates of sequence evolution may be enhanced in allopolyploid *Gossypium* relative to its diploid progenitors, although this pattern was difficult to statistically verify due to the relative recency of polyploid formation. Perhaps a more extensive sampling of genes would provide additional power to test the intriguing hypothesis that polyploidization leads to accelerated molecular evolutionary rates.

Materials and Methods

Genes Studied

The 48 genes selected (table 1) were from previous molecular phylogenetic studies and from ongoing investigations of genes expressed in developing cotton fibers. Many genes were chosen based on presumptive identifications after Blast searches of sequences derived from a *G. arboreum* cv. AKA8401 cotton fiber cDNA library (details of library at http://cfg.ucdavis.edu). Criteria for gene selection included likely importance in fiber development as well as copy number. We also included five *Adh* genes (Small et al. 2000), six genes encoding *R2R3-MYB* transcription factors (Loguercio, Zhang, and Wilkins 1999; Cedroni et al. 2002), two cellulose synthase genes (Cronn, Small, and Wendel 1999), and other putative single-copy genes (Cronn, Small, and Wendel 1999; Cronn et al. 2002*b*).

Copy number for all genes in both the diploids and the tetraploids was determined using standard methods of Southern hybridization (Brubaker and Wendel 1994; Cronn et al. 2002*b*; Small and Wendel 2000). These data (not shown) also facilitated selection of orthologous as opposed to paralogous genes, so that gene probes that yielded hybridization profiles indicative of complex multigene families were excluded from the study.

Plant Taxa

Species chosen for this study were based on the organismal framework provided in figure 1. *Gossypium raimondii* (unnamed accession) was chosen because it is the best living model of the D-genome donor (Wendel and Cronn 2003). The two A-genome diploids (*G. arboreum* and *G. herbaceum*) are phylogenetically equidistant from the A-genome progenitor of polyploid cotton and are thus interchangeable for the present purposes; in most cases *G. herbaceum* (GenBank accession numbers A1–73) was used, but *G. arboreum* (GenBank accession numbers A2–47) was substituted when amplification difficulties were

encountered. To represent allopolyploid cotton, either *G. barbadense* Pima S6 or *G. hirsutum* TM1 were used, depending on the bacterial artificial chromosomes (BAC) library used for sequence determination. To root phylogenetic trees and for purposes of providing reference sequences for relative rate tests, we included *Gossypioides kirkii* (Mast.) J. B. Hutchinson or *Kokia kauiensis* (Rock) O. Deg. and Duvel, representatives of sister genera that have been shown by phylogenetic analyses (Seelanan, Schnabel, and Wendel 1997; Wendel et al. 2002) to comprise the closest living relatives of *Gossypium* L. DNAs were isolated from young leaves using previously described protocols (Cedroni et al. 2002; Cronn, Small, and Wendel 1999; Cronn et al. 2002*b*; Wendel et al. 2002).

Amplification and Sequencing

Primers for PCR amplification and sequencing were designed as described elsewhere (Cedroni et al. 2002; Cronn, Small, and Wendel 1999; Small and Wendel 2000) or from cotton EST sequences in GenBank. A list of primers is available at http://www.botany.iastate.edu/~jfw/HomePage/jfwdata_sets.html. Two different amplification protocols were used on MJ Research thermocyclers. The first was a "touchdown PCR" method: 94°C for 3 min; 10 cycles of 94°C for 1 min, 56°C for 1 min, and 72°C for 2.5 min; 25 cycles of 94°C for 1 min, 56°C for 1 min, with a 0.6°C decrease per cycle, and 72°C for 2.5 min; 72°C for 7 min. Others were amplified using a hot-start method: 94°C for 3 min; 30 cycles of 94°C for 30 s, 54°C for 30 s, and 72°C for 1.25 min; 72°C for 6 min. Annealing temperature (48°C to 66°C) and extension time (1 to 3 min) requirements varied by gene, and these general conditions were adjusted on a gene-by-gene basis as necessary. Sequences that amplified with difficulty were cloned using standard T/A cloning protocols (pGem-T Easy [Promega Inc.]) and then sequenced from plasmids.

For each gene studied, allopolyploid species contain two homoeologous sequences, representing descendants of those contributed by the A-genome and D-genome donors at the time of polyploid formation. To isolate both homoeologs, we used one of two approaches. In one, heterogeneous PCR products were cloned after amplification from genomic DNA and the two duplicates were identified by restriction site analysis. Alternatively, homoeologs were isolated individually by PCR off of BAC clones derived from either *G. hirsutum* cv. Maxxa (Tomkins et al. 2001) or *G. barbadense* cv. Pima S6 (A. Paterson, unpublished data). Since each BAC clone contained only one of the two homoeologs, this latter strategy proved effective in minimizing problems of in vitro PCR recombination (Cronn et al. 2002*a*). BAC DNA was isolated from 50 ml cultures using the Psi-Clone Big BAC DNA Extraction Kit (Princeton Separations, Inc).

Automated sequencing was conducted using the ABI Big Dye v. 2.0 fluorescent primers and ABI Prism 377-3700 system at the Iowa State DNA Sequencing and Synthesis Facility. GenBank numbers for all sequences, aligned lengths for each gene, and putatively protein functions are listed in table 1.

Data Analysis

Sequences were aligned using BioEdit (Hall 1999) v. 5.0.9 (http://www.mbio.ncsu.edu/BioEdit/bioedit.html), and the resulting alignments were adjusted manually. DnaSP v. 3.53 (Rozas and Rozas 1999) was used to estimate G+C content and substitutions per site for synonymous (K_s), silent (K_{sil}, including both synonymous and noncoding sites), and replacement (K_a) sites. Phylogenetic analyses and Kimura two-parameter estimates of genetic distance were obtained using PAUP* (Swofford 1998). To evaluate possible rate heterogeneity among *Gossypium* lineages, relative rate tests were performed using the 1D tests of Tajima (1993). Analysis of variance was performed to determine whether lineage-specific estimates of divergence showed significant associations with genomes (A or D; fixed effect), ploidy levels (2X or 4X; fixed effect), or loci (random effect). For these analyses, evolutionarily inferred apomorphies for terminal lineages (A and D diploid, At and Dt tetraploid) were transformed into the number of inferred substitutions per kb of sequence. Subsequent generalized linear model analyses (PROC GLM in SAS v. 8.0; SAS Institute, Cary, N. C.) utilized untransformed values. Data sets that lacked outgroups (*A1834* and *AdhE*), included known or suspected pseudogenes (*AdhC*), or exhibited questionable orthology (*Myb4* and *E11*) were omitted from this analysis. For this reason, the number of genes included in analysis of variance was 43, rather than the 48 used in other computations.

Results

The genes studied encode a diverse set of proteins, including transcription factors, proteins presumed to be important in fiber growth and development (Wilkins and Jernstedt 1999), enzymes of intermediary metabolism, and others. Of the 48 genes surveyed, eight could not be assigned a function even though they showed high similarity to cDNA sequences described from other angiosperms. In these instances, exon and intron locations were inferred from alignments between genomic and cDNA sequences. The 48 genes had an average aligned length of approximately 850 base pairs (bp), with alignments ranging from 144 bp (partial sequence of *G8*) to 4,025 bp (complete *CesA1* sequence). For the five sequences evaluated per gene, an average of nine replacement substitutions occurred per alignment (range: 0 to 28), whereas the number of silent sites per alignment averaged 406 (range: 32 to 1,808). As expected, the average number of replacement substitutions between *Gossypium* sequences and orthologs from the outgroup ($K_a = 0.018$) is much lower than the number of silent substitutions ($K_{sil} = 0.061$), due to constraints at the amino acid level.

On average, G+C content was higher in coding (0.452 ± 0.045) than in noncoding regions (0.329 ± 0.039). G+C values for second (0.415 ± 0.078) and third (0.416 ± 0.078) codon positions were similar.

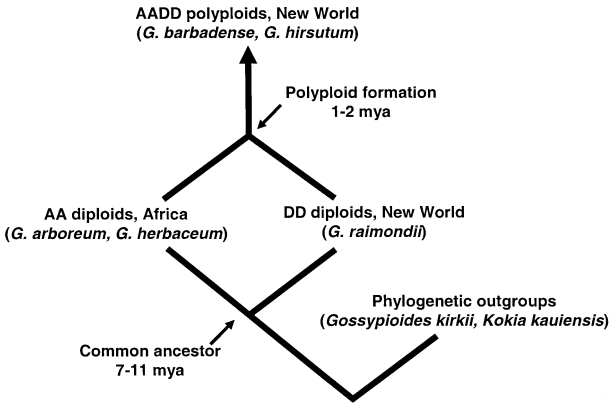

FIG. 1.—Phylogenetic framework for diploid and allopolyploid *Gossypium*, illustrating divergence of the genus from its closest relatives (the genera *Kokia* and *Gossypioides*), divergence of the A-genome and D-genome diploids from their most recent common ancestor, and the reunion of these two genomes upon polyploid (AD-genome) formation. Divergence dates estimated from previous studies (Wendel 1989; Seelanan, Schnabel, and Wendel 1997; Cronn et al. 2002*b*) are indicated, as are species used in the present study (in parentheses).

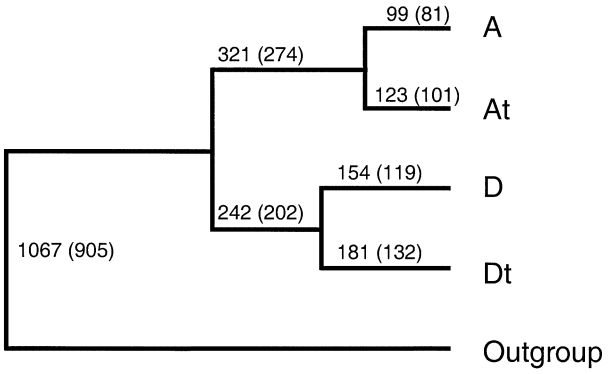

FIG. 2.—Global phylogeny estimated from summed data across 46 genes. Branch lengths represent total numbers of substitutions (both silent and replacement), and numbers in parentheses show the number of silent substitutions. Two genes (*A1834* and *AdhE*) were excluded because outgroup sequences were unavailable.

Gene Tree Topologies

Because diploid cottons diverged from each other recently relative to the scale of molecular evolution for single-copy nuclear genes, homoplasy was expected to be low. This was indeed the case for all 48 genes, with consistency indices ranging from 0.92 to 1.00, with a mean of 0.99 across the 48 gene trees. Accordingly, parsimony analysis yielded unambiguously resolved topologies that are congruent with the well-documented (Cronn et al. 2002*b*; Wendel and Cronn 2003) phylogeny of *Gossypium* shown in figure 1. One test provided by these 48 gene trees is that of gene conversion or recombination between homoeologs, which would be expected to lead either to elevated homoplasy or an altered phylogenetic topology. As shown previously for a smaller set of genes (Cronn, Small, and Wendel 1999), there was no evidence in the 48 genes studied here for these forms of interlocus interaction.

A composite phylogenetic tree with summed branch lengths representing the total number of inferred substitutions (silent and replacement) is shown in figure 2. Summed across branches, 816 substitutions are inferred to have occurred in the gene regions sampled since divergence of the A and D diploids from their common ancestor, with a slight acceleration in the polyploid genomes (total = 867). Similarly, in both the A-genome and the D-genome branches, more substitutions were observed in the polyploid genome than in orthologs from the corresponding diploid (123 versus 99 for the A-genome; 181 versus 154 for the D-genome). As expected from previous data (reviewed in Wendel and Cronn 2003), *G. herbaceum* and *G. arboreum* are closer models of the A-genome ancestor than *G. raimondii* is of the D-genome donor; viz., branch lengths are shorter in the A-genome (total of 222 substitutions) than in the D-genome (335 substitutions) clade (fig. 2).

Rate Variation Among Nuclear Genes

For each gene, we calculated Kimura two-parameter distances (K2P) between the two genomes of *Gossypium* polyploids (At and Dt), between the diploid progenitor genomes (A and D); between each diploid and its counterpart in allopolyploid cotton (A and At, D and Dt), and between all species and the outgroup (table 2). For the A-D comparison, the lowest K2P value was 0.003 (locus *G8*), whereas the highest estimate was 0.051 (locus *E11*), with a mean of 0.022 and a median of 0.021. For the At-Dt comparisons, the lowest K2P value was 0.008 (locus *A1520*) and the highest was 0.054 (locus *C7*), with a mean of 0.024 and a median of 0.023.

Because agronomic selection may have altered molecular evolutionary patterns for genes involved in fiber development, we parsed the data into "fiber" and "nonfiber" genes. In all interspecific comparisons and for all genes (data not shown), including between wild and domesticated species, K2P values for nonfiber genes (n = 12) were nearly identical to those obtained for genes expressed in fibers (n = 36).

To explore in more detail the nature of substitutions contributing to overall divergence in *Gossypium*, we tabulated levels of synonymous (K_s), replacement (K_a), and silent (K_{sil}) substitutions and K_a:K_{sil} ratios (table 3). As expected from the organismal phylogeny, comparisons between *Gossypium* and the phylogenetic outgroup yielded the highest K_s and K_a values in all cases (see also table 2). Average K_s and K_a values are higher in the A-D and At-Dt comparisons than in the A-At and D-Dt comparisons, as we would expect because the A-D divergence occurred well before the formation of the polyploids. Also, sequence divergence was lower in A-At comparisons than in D-Dt, as noted above and as evidenced by the summary in figure 2. Sequence divergence amounts between the parental species (A and D) and their respective genomes in the polyploid (At and Dt) are not evidently different, although this is not easily statistically tested since so much of their evolutionary history is shared (the divergence estimates are not independent). Replacement substitutions are lower in all

Table 2
Kimura Two-Parameter Distances for 48 *Gossypium* Genes

Comparison[a]	K2P +/− Standard Error (Range)
A vs. D	0.022 +/− 0.011 (0.003–0.048)
A vs. At	0.007 +/− 0.008 (0.0–0.040)
D vs. Dt	0.010 +/− 0.007 (0.0–0.027)
At vs. Dt	0.024 +/− 0.010 (0.004–0.054)
A vs. outgroup	0.040 +/− 0.016 (0.016–0.107)
At vs. outgroup	0.042 +/− 0.017 (0.016–0.106)
D vs. outgroup	0.040 +/− 0.016 (0.011–0.101)
Dt vs. outgroup	0.040 +/− 0.016 (0.016–0.094)
Total aligned length	40,685

[a] Taxon codes are as follows: *Gossypium arboreum* or *G. herbaceum* (A), *G. raimondii* (D), *G. barbadense* or *G. hirsutum* "A homoelog" (At), *G. barbadense* or *G. hirsutum* "D homoeolog" (Dt), and *Gossypioides kirkii* or *Kokia kauiensis* (outgroup).

pairings compared with corresponding K_s and K_{sil} estimates.

Substitution ratios, specifically K_a:K_{sil}, can be informative with respect to the strength and direction of selection. If K_a:$K_s < 1$, then purifying selection is implicated, presumably due to constraints at the level of amino acid substitutions. If, however, K_a:$K_s > 1$, then positive selection may be involved (Yang and Bielawski 2000). Results from table 3 yield no evidence of positive selection, as K_a:K_{sil} values are all relatively low. K_a:K_{sil} ratios for the A-At and D-Dt comparisons are higher than for the A-D and At-Dt comparisons, but this likely is due to stochastically large errors on ratios of small numbers rather than selection. If selection was acting at the tetraploid level, for example, due to functional diversification of duplicated genes, then the At-Dt value should be similarly high, but it is not.

For each gene, estimates of K2P were calculated between the phylogenetic outgroup and each *Gossypium* genome. Divergences between each ingroup sequence and that of the outgroup were similar in all cases, as expected under conditions of rate homogeneity and the phylogeny provided in figure 1. This comparison provides an informal relative rate test, suggesting that there is no rate variation among the *Gossypium* taxa studied (as indicated also by formal relative rate tests; see below).

Given the range in molecular evolutionary rates among genes studied, it was of interest to explore the distribution of rates among genes and estimate a mean rate for both diploid and allopolyploid cotton. K2P values for all 48 genes visually appeared to be approximately normally distributed for both the A-D and the At-Dt comparisons (not shown), but Shapiro-Wilk tests of normality were marginally rejected at the 0.05 level. Mean divergences for diploid versus polyploid sequences were similar (0.0223 ± 0.0015 versus 0.0236 ± 0.0014), as were standard deviations around these means (0.011 versus 0.010).

Rate Variation Among *Gossypium* Lineages

To explore whether rate heterogeneity existed among *Gossypium* lineages for any of the 48 genes, we used the Tajima relative rate test (Tajima 1993). In nearly all cases, the resulting values were not significant, indicating approximate rate equivalence among lineages for each gene. Only four instances of significant ($0.5 < P < 0.01$) rate heterogeneity were indicated among the nearly 200 tests, as expected by chance alone. These were accelerated rates for the At homoeolog of *C7*, the D ortholog of *G3*, the A ortholog of *CLK1*, and the D ortholog of *GhMYB4*. Of these exceptions, *C7* is perhaps the most interesting due to the number of replacements (28 changes in 839 nt) and the relatively high K_a:K_{sil} ratio (table 1). We note that the two branches leading to A/At and D/Dt were long considering the aligned length, suggesting that either *C7* is a relatively fast-evolving gene or that we inadvertently isolated paralogs from one of these two clades.

One implication of rate homogeneity is that the evolutionary rate for each gene in table 1 is a property of that gene rather than the particular *Gossypium* lineage to which it belongs. To examine this suggestion further, we calculated the correlation between synonymous substitution rates in A-At and D-Dt comparisons. Since these are independent divergences, a high correlation is expected only if molecular evolutionary rate reflects inherent properties of the gene and/or its genomic context. Pearson's correlation coefficient for these two sets of divergences, using K_s for the entire gene sequence (or just the coding region when only exons were available) was calculated to be 0.98, indicating a strong correlation.

Although relative rate tests and mean K2P divergence levels indicated that orthologous genes in diploid *Gossypium* accumulate nucleotide substitutions at approximately the same rate as their counterparts in the allopolyploids, the branch lengths of figure 2 suggest that there may be a slight elevation in molecular evolutionary rate in the allopolyploid. As noted above, in both the A-genome and D-genome branches, more substitutions were observed in the polyploid genome (At and Dt) than

Table 3
Molecular Evolutionary Rates in Various Comparisons Among *Gossypium* Taxa

Rate	A-D	A-At	D-Dt	At-Dt	Ingroup-Outgroup
K_s	0.038 +/− 0.036	0.009 +/− 0.017	0.011 +/− 0.017	0.042 +/− 0.025	0.074 +/− 0.045
K_a	0.009 +/− 0.014	0.003 +/− 0.006	0.005 +/− 0.011	0.010 +/− 0.013	0.016 +/− 0.014
K_{sil}	0.031 +/− 0.014	0.007 +/− 0.007	0.011 +/− 0.007	0.032 +/− 0.017	0.059 +/− 0.023
K_a:K_{sil}	0.290	0.429	0.455	0.313	0.271

NOTE.—K_s, K_a, K_{sil}, and K_a:K_{sil} denote weighted (by nucleotides per gene) mean substitutions at synonymous sites, nonsynonymous sites, synonymous sites within codons plus all noncoding positions, and ratio or replacement to silent sites, respectively. Taxon codes are as follows: *Gossypium arboreum* or *G. herbaceum* (A), *G. raimondii* (D), *G. barbadense* or *G. hirsutum* "A homoelog" (At), *G. barbadense* or *G. hirsutum* "D homoeolog" (Dt), and *Gossypioides kirkii* or *Kokia kauiensis* (outgroup). Ingroup-outgroup values are means among the pairwise comparisons between *Gossypium* and outgroup orthologs. $N = 48$.

in orthologs from the corresponding diploid (123 versus 99 for the A-genome;181 versus 154 for the D-genome). An analysis of variance showed both a locus ($P = 0.003$) and a genome (D faster than A; $P = 0.001$) effect but failed to detect a significant effect due to ploidy level ($P = 0.16$). However, when the data were tabulated on a gene-by-gene basis, there were 41 instances in which the number of substitutions was higher in the branch leading to the polyploid than to the diploid genome, whereas the reverse was true in only 26 cases (in 31 cases they were equal). Under the hypothesis of equal rates in polyploids and diploids, this difference may be interpreted as marginally significant (41 versus 26 with the expectation that these numbers would be equal [$\chi^2 = 3.36$; $0.10 > P > 0.05$]).

Discussion

Rate Variation Among *Gossypium* Nuclear Genes

Inspection of the levels of silent substitution for 48 single-copy nuclear genes in *Gossypium* leads to several observations and conclusions. First, the evolutionary rate for a given gene appears largely to be a property of that gene and/or its genomic milieu. This is evidenced by the near-perfect correlation between silent site divergences among the two independent lineage-pairs, A to At and D to Dt, as well as the observation that rate heterogeneity is rare among lineages (Tajima relative rate tests). The idea that there exist genes and/or genomic regions that are more variable than others is substantiated by a wealth of studies in a diverse array of organisms. Possible explanations for silent site rate variation among genes within an organism include genomic location and possible correlates with local levels of recombination and mutation (Begun and Aquadro 1992; Stephan and Langley 1998; Williams and Hurst 2000; Lercher and Hurst 2002; Zhang, Vision, and Gaut 2002), variation in GC content (Ticher and Grauer 1989), variable codon usage biases (Sorhannus and Fox 1999; Tiffin and Hahn 2002), and perhaps poorly understood differences in chromatin structure that alter in unknown ways the rate of fixation of mutations.

A second noteworthy aspect of the *Gossypium* data is that silent substitution rates vary widely among genes. The range of silent substitution amounts among the 48 genes varies ninefold, from 0.018 to 0.162, although if the highest two and lowest two values are excluded, the range narrows considerably, from 0.035 to 0.101 (table 1). Thus, 90% of the values fall within a threefold range, which is remarkably similar to the 2.6-fold range recently reported for 242 gene pairs in *Arabidopsis* (Zhang, Vision, and Gaut 2002). Our data contribute to and exemplify the increasingly common reports of rate variation in a diverse assemblage of organisms (Wolfe, Sharp, and Li 1989; Moriyama and Gojobori 1992; Wolfe and Sharp 1993; Collins and Jukes 1994; Moriyama and Powell 1996; Zeng et al. 1998; Kusumi et al. 2002; Tiffin and Hahn 2002; Zhang, Vision, and Gaut 2002). These examples, drawn from across the phylogenetic spectrum, illustrate the generality that intergenic rate variation is a fundamental feature of complex eukaryotic genomes.

A third notable aspect of the *Gossypium* data concerns the ratio of replacement to silent substitutions. Mean K_a:K_{sil} ratios range from 0.271 to 0.455 for various intergenomic comparisons (table 3), with the lowest ratio corresponding to the greatest evolutionary distance (ingroup-outgroup) and the highest ratios corresponding to the smallest evolutionary distance (A-At and D-Dt). This suggestion of a relationship between degree of divergence and K_a:K_{sil} ratios is consistent with the notion that some proportion of amino acid substitutions are relatively neutral, vis-à-vis protein function, and hence nucleotide substitutions that cause these changes are neutral or near-neutral. These sites might be expected to behave more like silent sites in terms of evolutionary rates, but they also would become saturated more rapidly than other replacement sites. Hence, as evolutionary distance increases, the near-neutral replacement sites approach saturation, but purifying selection on nonneutral sites continues to retard accumulation of replacements. The net effect is that K_a:K_{sil} ratios become smaller as evolutionary distance increases (or, alternatively, that K_{sil}:K_a ratios become larger). This proposal is consistent with other studies involving more divergent taxa than those studied here, although the correlation with time may not be particularly tight due to the numerous other features that potentially influence molecular evolutionary rate. For anciently duplicated *Arabidopsis* genes, for example, the K_s:K_a ratio is 5 (Zhang, Vision, and Gaut 2002), whereas for rat and mouse the ratio is approximately 7 (Wolfe and Sharp 1993). Among genera of Cupressaceae, which may have diverged approximately 100 MYA, silent substitutions outnumber replacements by sevenfold to eightfold (Kusumi et al. 2002), with a similar ratio for 218 orthologs from *Arabidopsis-Brassica*, representing perhaps 35 Myr of divergence (Tiffin and Hahn 2002).

Phylogenetic Analysis and Independent Evolution of Homoeologs

Parsimony analysis for each of the 48 genes led to the recovery of the topology expected from our phylogenetic understanding of *Gossypium* (fig. 2). These results confirm and extend those of an earlier analysis of duplicate gene evolution in allopolyploid *Gossypium* (Cronn, Small, and Wendel 1999), which demonstrated that homoeologs in polyploid cotton evolve independently of one another in the allopolyploid nucleus. This stands in contrast to at least some repetitive DNAs, which experience postpolyploidization interlocus homogenization (Wendel et al. 1995) as a consequence of one or more processes of concerted evolution. Our results lend additional weight to the suggestion that intergenomic interactions between duplicated single-copy genes is uncommon in allopolyploids (Cronn, Small, and Wendel 1999; Wendel 2000). A similar conclusion was recently reached for 242 gene pairs duplicated by ancient polyploidy in *Arabidopsis*, where tests failed to provide evidence of gene conversion for any of the duplicates (Zhang, Vision, and Gaut 2002).

Rates of Gene Evolution in Diploid and Allopolyploid *Gossypium*

As shown in table 2, the mean K2P divergence of orthologs between A-genome and D-genome diploid

cottons for 48 single-copy nuclear genes is 0.022, a value nearly identical to that obtained for the same genes isolated from the two descendent genomes (At and Dt) in allopolyploid cotton (0.024). Also, analysis of variance failed to detect an effect of ploidy on rates of sequence evolution. Thus, based on 40,000 nucleotides per taxon, these analyses suggest equivalent rates of genic evolution in diploid and allopolyploid cotton. We note, however, that the total number of substitutions in branches leading to the polyploid genomes (At and Dt) are approximately 20% higher than those of their diploid counterparts (fig. 2) and that a marginally significant ($0.10 > P > 0.05$) effect of polyploidy on nucleotide substitutions was revealed by analysis of branch lengths on a gene-by-gene basis. These data lead to the suggestion that polyploidy in *Gossypium* has been accompanied by a modest rate enhancement, as also suggested in an earlier study (Cronn, Small, and Wendel 1999). Perhaps it is not surprising that the rate acceleration is difficult to detect statistically, as polyploid *Gossypium* formed relatively recently, and so branch lengths are small and subject to proportionately high stochastic variation.

Although there exist no comparable surveys of homoeologous gene evolution in other plant polyploids, a number of studies have demonstrated dramatic genetic and epigenetic changes immediately after polyploidy in some plant groups (Song et al. 1995; Feldman et al. 1997; Liu, Vega, and Feldman 1998; Liu et al. 1998; Comai et al. 2000; Ozkan, Levy, and Feldman 2001; Shaked et al. 2001; Kashkush, Feldman, and Levy 2002). The present study complements other recent analyses of polyploid genome evolution in *Gossypium* (Brubaker, Paterson, and Wendel 1999; Cronn, Small, and Wendel 1999; Liu et al. 2001) in showing that polyploidy is not accompanied by rapid genome change. In this context we detected no cases of gene loss or gene conversion, and rate evolutionary rate enhancements, if real, are only modest. Similar relative genomic stasis has been reported for the young allopolyploid grass, *Spartina anglica* (Baumel, Ainouche, and Levasseur 2001), and *Brassica juncea* (*contra* Song et al. 1995; Axelsson et al. 2000). As suggested here for *Gossypium*, it will be of interest to explore whether silent substitution rates for single-copy nuclear genes are elevated relative to their diploid progenitors in allopolyploid systems that are subject to genomic instability, such as in *Aegilops/Triticum* and *Brassica*.

Modern Diploids and the Ancestors of Polyploid Cotton

Ever since the discovery that tetraploid *Gossypium* species contain two different genomes, investigators have attempted to address the question of parentage; that is, which of the modern species of A-genome and D-genome diploids best serve as models of the progenitor genome donors? Over the decades, a diverse array of tools have addressed this question (reviewed in Wendel and Cronn 2003), collectively demonstrating that the best extant models of the ancestral genome donors are *G. arboreum* and *G. herbaceum* (A-genome) and *G. raimondii* (D-genome). Cytogenetic and segregation data suggested that the A-genome of allopolyploid cotton is more similar to that of the A-genome diploids than the D-genome of the allopolyploid is to that of the D-genome diploids. For example, in synthetic allohexaploids formed between diploid and allopolyploid cotton, multivalent frequencies are higher and genetic segregation more closely approximates autotetraploid ratios for A-genome chromosomes than for D-genome chromosomes (Gerstel and Phillips 1958; Phillips 1964). Subsequent data from many sources has confirmed this observation (Wendel and Cronn 2003). Cronn, Small, and Wendel (1999) quantified these relationships using 14,705 nt of sequence information for 16 nuclear loci isolated from the D-genome diploid *G. raimondii*, the A-genome diploid *G. arboreum* (or *G. herbaceum*), and the AD-genome tetraploid *G. hirsutum*, much as in the present study but with a smaller sampling of genes. Sequence divergence between the diploids and their corresponding genomes in the allopolyploid were 0.68% and 1.05% for the A-genomes and D-genomes, respectively. The present study confirms and extends this understanding: Kimura two-parameter genetic distances between A and At and between D and Dt are 0.007 and 0.010, respectively (table 2), and this same quantitative relationship is captured in the branch lengths of figure 2 (222 versus 335 total substitutions distinguishing the two genomes in the A and D clades, respectively). Thus, *G. arboreum* and *G. herbaceum* may be thought of as an approximately 50% better model of the progenitor A-genome diploid than *G. raimondii* is of the D-genome diploid.

Gene Evolution, the Molecular Clock, and the Age of Polyploidy in *Gossypium*

Abundant evidence establishes that the five species of tetraploid cottons are allopolyploids containing one genome similar to those found in the Old World, A-genome diploids, and a second genome like those of the New World, D-genome diploids (reviewed in Wendel and Cronn 2003). Because the two parental genome groups exist in diploid species that presently occupy different hemispheres, the question of how and when allopolyploid cotton formed has stimulated discussion for more than 50 years. Some authors have suggested that *Gossypium* had an ancient, perhaps Cretaceous origin, due to its global distribution and high level of cytogenetic and morphological diversity, whereas others have invoked a origin of allopolyploids in agricultural times, forwarding a scenario that involved human transfer of an African or Asiatic A-genome cultigen to the New World, followed by deliberate or accidental hybridization with a wild D-genome species. These speculations and others, which encompass proposals ranging from a Cretaceous (60 to 100 MYA) to a recent (6,000 years ago) origin, are discussed at length in Wendel and Cronn (2003).

DNA sequence data have uniformly supported the view that allopolyploid *Gossypium* originated prior to the evolution of modern humans but relatively recently in geological terms, perhaps during the Pleistocene 1 to 2 MYA (Wendel 1989; Seelanan, Schnabel, and Wendel 1997; Small et al. 1998; Cronn, Small, and Wendel 1999). Cronn, Small, and Wendel (1999), in a study of 16 low-

copy nuclear sequences, reported that mean sequence divergence between the diploids and their counterparts in the allopolyploid averaged 0.68% and 1.05%, respectively, for the A-genome and D-genome comparisons. Similar values (K_{sil} = 0.7% and 1.1%, respectively; K_s = 0.9% and 1.1%, respectively [table 3]) were obtained in the present study using a mean rate based on three times as many genes. Relative rate tests revealed no evidence of lineage-specific effects, so this source of potential error in molecular clock applications is minimized.

This leaves clock calibration as the most troublesome source of error in estimating divergence dates. Although rates of synonymous site evolution have been estimated for several plants using a small sample of genes (2.6×10^{-9} – 1.5×10^{-8} substitutions/synonymous site/year [Morton, Gaut, and Clegg 1996; Gaut 1998; Koch, Haubold, and Mitchell-Olds 2000]), little is known about the general utility of these estimates. To the extent that they are applicable, and given that generation time is negatively correlated with molecular evolutionary rates (Gaut 1998) and that wild *Gossypium* species are long-lived perennials, it is likely that the more appropriate end of the spectrum to use is the slower rates. To estimate the age of allopolyploid formation, the values for the A-genome listed above are the most relevant, noting that these data will provide the *maximum* age of *Gossypium* allopolyploids. This is because modern A-genome diploid cottons may not be the *direct* descendants of the actual genome donors. Instead, we only know that they are the closest living model of the ancestral diploid implicated in allopolyploid formation.

Using the formula $T = K/2r$, where K equals divergence amount (K_{sil}, K_s from table 3) and r corresponds to the rate of divergence for nuclear genes from plants (2.6×10^{-9} substitutions/site/year), we estimate that allopolyploids formed 1.3 to 1.7 MYA, depending on whether silent (synonymous plus non-coding) or just synonymous sites are used in the calculation. Hence, it seems probable that *Gossypium* allopolyploids formed in the Mid-Pleistocene, circa 1 to 2 MYA, as suggested by other authors using different criteria (Wendel and Cronn 2003). Extending the analysis to the diploids, we estimate that the A-genome and D-genome lineages diverged from one another 6.0 to 7.3 MYA and that *Gossypium* last shared a common ancestor with its closest relatives (*Gossypioides* and *Kokia*) 11.3 to 14.2 MYA. Thus, the two diploid genomes, A and D, experienced approximately 5 Myr of evolution in isolation from one another prior to their reunion at the time of polyploid formation during the Pleistocene.

Several sources of error remain unaccounted for in the foregoing calculations, and clearly clock calibration remains an important consideration. For example, if the less likely (in our opinion) faster rate estimates reported above (1.5×10^{-8} substitutions/synonymous site/year) are used, polyploid formation may be estimated to have occurred as recently as 230,000 to 300,000 years ago. Using 48 genes, however, and establishing rate homogeneity in the taxa under study, lends a degree of confidence to the interpretation offered that polyploid formation is of mid-Pleistocene age. Additional insight into the accuracy of this inference will require more data on absolute rates of synonymous site divergence, which remain the largest single source of possible error.

Acknowledgments

We are grateful to Kara Shockey for *CLK1* sequences, Corrinne Grover for *CesA* sequence data, Ryan Percifield for technical assistance, and Keith Adams for discussion. Financial support was provided by the National Science Foundation, the US-Israel Binational Science Foundation, the Plant Sciences Institute of Iowa State University, and the Spanish Ministry of Education, Culture and Sports.

Literature Cited

Axelsson, T., C. M. Bowman, A. G. Sharpe, D. J. Lydiate, and U. Lagercrantz. 2000. Amphidiploid *Brassica juncea* contains conserved progenitor genomes. Genome **43**:679–688.

Baumel, A., M. L. Ainouche, and J. E. Levasseur. 2001. Molecular investigations in populations of *Spartina anglica* C. E. Hubbard (Poaceae) invading coastal Brittany (France). Mol. Ecol. **10**:1689–1701.

Begun, D. J., and C. F. Aquadro. 1992. Levels of naturally occurring DNA polymorphism correlate with recombination rates in *D. melanogaster*. Nature **356**:519–520.

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). Am. J. Bot. **81**:1309–1326.

Cedroni, M. L., R. C. Cronn, K. L. Adams, T. A., Wilkins, and J. F. Wendel. 2002. Evolution and expression of *MYB* genes in diploid and polyploid cotton. Plant Mol. Biol. **51**:313–325.

Collins, D. W., and T. H. Jukes. 1994. Rates of transition and transversion in coding sequences since the human-rodent divergence. Genomics **20**:386–396.

Comai, L., A. P. Tyagi, K. Winter, R. Holmes-Davis, S. H. Reynolds, Y. Stevens, and B. Byers. 2000. Phenotypic instability and rapid gene silencing in newly formed *Arabidopsis* allotetraploids. Plant Cell **12**:1551–1567.

Cronn, R. C., M. Cedroni, T. Haselkorn, C. Osborne, and J. F. Wendel. 2002*a*. PCR-mediated recombination in amplification products derived from polyploid cotton. Theor. Appl. Genet. **104**:482–489.

Cronn, R. C., R. L. Small, T. Haselkorn, and J. F. Wendel. 2002*b*. Rapid diversification of the cotton genus (*Gossypium*: Malvaceae) revealed by analysis of sixteen nuclear and chloroplast genes. Am. J. Bot. **89**:707–725.

Cronn, R. C., R. L. Small, and J. F. Wendel. 1999. Duplicated genes evolve independently after polyploid formation in cotton. Proc. Natl. Acad. Sci. USA **96**:14406–14411.

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

Endrizzi, J. E., F. R. H. Katterman, and R. F. Geever. 1989. DNA hybridization and time of origin of three species of *Gossypium*. Evol. Trends Plants **3**:553–559.

Feldman, M., B. Liu, G. Segal, S. Abbo, A. A. Levy, and J. M. Vega. 1997. Rapid elimination of low-copy DNA sequences in polyploid wheat: a possible mechanism for differentiation of homoeologous chromosomes. Genetics **147**:1381–1387.

Galau, G. A., and T. A. Wilkins. 1989. Alloplasmic male sterility in AD allotetraploid *Gossypium hirsutum* upon replacement of its resident A cytoplasm with that of the D species *G. harknessii*. Theor. Appl. Genet. **78**:23–30.

Gaut, B. S. 1998. Molecular clocks and nucleotide substitution rates in higher plants. Pp. 93–120 *in* M. K. Hecht, ed. Evolutionary biology. Plenum Press, New York.

Gerstel, D. U., and L. L. Phillips. 1958. Segregation of synthetic amphiploids in *Gossypium* and *Nicotiana*. Cold Spring Harbor Symp. Quant. Biol. **23**:225–237.

Hall, T. A. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp. Ser. **41**:95–98.

Hillis, D. M., C. Moritz, and B. K. Mable. 1996. Molecular systematics. 2nd edition. Sinauer Associates, Sunderland, Mass.

Kashkush, K., M. Feldman, and A. A. Levy. 2002. Gene loss, silencing, and activation in a newly synthesized wheat allotetraploid. Genetics **160**:1651–1659.

Koch, M. A., B. Haubold, and T. Mitchell-Olds. 2000. Comparative evolutionary analysis of chalcone synthase and alcohol dehydrogenase in *Arabidopsis*, *Arabis*, and related genera (Brassicaceae). Mol. Biol. Evol. **17**:1483–1498.

Kusumi, J., Y. Tsumura, H. Yoshimaru, and H. Tachida. 2002. Molecular evolution of nuclear genes in Cupressaceae, a group of conifer trees. Mol. Biol. Evol. **19**:736–747.

Lercher, M. J., and L. D. Hurst. 2002. Human SNP variability and mutation rate are higher in regions of high recombination. Trends Genet. **18**:337–340.

Liu, B., C. L. Brubaker, G. Mergeai, R. C. Cronn, and J. F. Wendel. 2001. Polyploid formation in cotton is not accompanied by rapid genomic changes. Genome **44**:321–330.

Liu, B., J. M. Vega, and M. Feldman. 1998. Rapid genomic changes in newly synthesized amphiploids of *Triticum* and *Aegilops*. II. Changes in low-copy coding DNA sequences. Genome **41**:535–542.

Liu, B., J. M. Vega, G. Segal, S. Abbo, M. Rodova, and M. Feldman. 1998. Rapid genomic changes in newly synthesized amphiploids of *Triticum* and *Aegilops*. I. Changes in low-copy noncoding DNA sequences. Genome **41**:272–277.

Loguercio, L. L., J.-Q. Zhang, and T. A. Wilkins. 1999. Differential regulation of six novel MYB-domain genes defines two distinct expression patterns in allotetraploid cotton (*Gossypium hirsutum* L.). Mol. Gen. Genet. **261**:660–671.

Moriyama, E. N., and J. R. Powell. 1996. Intraspecific nuclear DNA variation in *Drosophila*. Mol. Biol. Evol. **13**:261–277.

Morton, B. R., B. S. Gaut, and M. T. Clegg. 1996. Evolution of alcohol dehydrogenase genes in the palm and grass families. Proc. Natl. Acad. Sci. USA **93**:11735–11739.

Ozkan, H., A. A. Levy, and M. Feldman. 2001. Allopolyploidy-induced rapid genome evolution in the wheat (*Aegilops-Triticum*) group. Plant Cell **13**:1735–1747.

Phillips, L. L. 1964. Segregation in new allopolyploids of *Gossypium*. V. Multivalent formation in New World × Asiatic and New World × wild American hexaploids. Am. J. Bot. **51**:324–329.

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

Sanderson, M. 1998. Estimating rate and time in molecular phylogenies: beyond the molecular clock? Pp. 242–264 *in* P. S. Soltis, D. E. Soltis, and J. J. Doyle, eds. Molecular systematics of plants. Kluwer, Boston.

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

Shaked, H., K. Kashkush, H. Ozkan, M. Feldman, and A. A. Levy. 2001. Sequence elimination and cytosine methylation are rapid and reproducible responses of the genome to wide hybridization and allopolyploidy in wheat. Plant Cell **13**:1749–1759.

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 group. Am. J. Bot. **85**:1301–1315.

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, P. S., D. E. Soltis, V. Savolainen, P. R. Crane, and T. G. Barraclough. 2002. Rate heterogeneity among lineages of tracheophytes: integration of molecular and fossil data and evidence for molecular living fossils. Proc. Natl. Acad. Sci. USA **99**:4430–4435.

Song, K., P. Lu, K. Tang, and T. C. Osborn. 1995. Rapid genome change in synthetic polyploids of *Brassica* and its implications for polyploid evolution. Proc. Natl. Acad. Sci. USA **92**:7719–7723.

Sorhannus, U., and M. Fox. 1999. Synonymous and nonsynonymous substitution rates in diatoms: a comparison between chloroplast and nuclear genes. J Mol Evol **48**:209–212.

Stephan, W., and C. H. Langley. 1998. DNA polymorphism in *Lycopersicon* and crossing-over per physical length. Genetics **150**:1585–1593.

Swofford, D. L. 1998. PAUP*: phylogenetic analysis using parsimony (*and other methods). Version 4.0. Sinauer Associates, Sunderland, Mass.

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

Ticher, A., and D. Grauer. 1989. Nucleic acid composition, codon usage, and the rate of synonymous substitution in protein-coding genes. J. Mol. Evol. **28**:286–298.

Tiffin, P., and M. W. Hahn. 2002. Coding sequence divergence between two closely related plant species: *Arabidopsis thaliana* and *Brassica rapa* ssp. *pekinensis*. J. Mol. Evol. **54**:746–753.

Tomkins, J. P., D. G. Peterson, T. J. Yang, D. Main, T. A. Wilkins, A. H. Paterson, and R. A. Wing. 2001. Development of genomic resources for cotton (*Gossypium hirsutum* L.): BAC library construction, preliminary STC analysis, and identification of clones associated with fiber development. Mol. Breed. **8**:255–261.

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

Wendel, J. F. 2000. Genome evolution in polyploids. Plant Mol. Biol. **42**:225–249.

Wendel, J. F., and R. C. Cronn. 2003. Polyploidy and the evolutionary history of cotton. Adv. Agron. **78**:139–186.

Wendel, J. F., R. C. Cronn, J. S., Johnston, and H. J. Price. 2002. Feast and famine in plant genomes. Genetica **115**:36–47.

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

Wilkins, T. A., and J. A. Jernstedt. 1999. Molecular genetics of developing cotton fibers. Pp. 231–267 *in* A. S. Basra, ed. Cotton fibers. Haworth Press, New York.

Williams, E. J., and L. D. Hurst. 2000. The proteins of linked genes evolve at similar rates. Nature **407**:900–903.

Wolfe, K. H., and P. M. Sharp. 1993. Mammalian gene evolution: nucleotide sequence divergence between mouse and rat. J. Mol. Evol. **37**:441–456.

Wolfe, K. H., P. M. Sharp, and W.-H. Li. 1989. Rates of

synonymous substitution in plant nuclear genes. J. Mol. Evol. **29**:208–211.

Yang, Z., and J. P. Bielawski. 2000. Statistical methods for detecting molecular adaptation. Trends Ecol. Evol. **15**:496–503.

Zeng, L.-W., J. M. Comeron, B. Chen, and M. Kreitman. 1998. The molecular clock revisited: the rate of synonymous versus replacement change in *Drosophila*. Genetica **102/103**:369–382.

Zhang, L., T. J. Vision, and B. S. Gaut. 2002. Patterns of nucleotide substitution among simultaneously duplicated gene pairs in *Arabidopsis thaliana*. Mol. Biol. Evol. **19**: 1464–1473.

Brandon Gaut, Associate Editor

Accepted December 5, 2002