Letter to the Editor

Intron Size and Genome Size in Plants

Jonathan F. Wendel, Richard C. Cronn,*† Ines Alvarez,* Bao Liu,*‡ Randall L. Small,*§ and David S. Senchina**

*Department of Botany, Iowa State University; †Pacific Northwest Research Station, USDA Forest Service, Corvallis; ‡Institute of Genetics and Cytology, Northeast Normal University, Changchun; and §Department of Botany, University of Tennessee, Knoxville

It has long been known that genomes vary over a remarkable range of sizes in both plants (Bennett, Cox, and Leitch 1997) and animals (Gregory 2001). It also has become evident that across the broad phylogenetic sweep, genome size may be correlated with intron size (Deutsch and Long 1999; Vinogradov 1999; McLysaght et al. 2000), suggesting that some component of genome size evolution takes place within genes. Examples include humans and pufferfish (*Fugu*), where comparisons of 199 introns in 22 orthologous genes showed that introns in *Fugu* were on average eight times as small as those in humans, consistent with their ratio of genome sizes (McLysaght et al. 2000). Similarly, Deutsch and Long (1999) tabulated intron sizes across a broad phylogenetic spectrum of eukaryotes and noted a general but weak correlation with genome size, with humans having the most and longest introns (mean of 3.4 kbp) among the 10 taxa studied. Intron size is also correlated with genome size in *Drosophila* (Moriyama, Petrov, and Hartl 1998), showing that the correlation may extend to more recent divergences.

At present there is little information on the correlation between genome and intron sizes in plants, although there are suggestions that plants with small genomes have smaller introns (Deutsch and Long 1999; Vinogradov 1999). Whereas broad comparisons across widely divergent taxa are now possible given completed draft sequences for the rice (Goff et al. 2002; Yu et al. 2002) and *Arabidopsis* (Arabidopsis Genome Initiative 2000) genomes, the divergence time between Poaceae and Brassicaceae is so large that the influence of genome size on intron size may be confounded by numerous other, unstudied covariables. It seems likely that more informative studies will involve closely related taxa that vary significantly in genome size but which share recent evolutionary history and a broad suite of life-history features. An additional advantage of comparing close relatives is that orthology among genes, and hence introns, may be more readily established. This latter point may be especially important, given the relative lability of copy-number for many gene families (Small and Wendel 2000).

To exemplify this approach, we studied the relationship between intron size and genome size for orthologous genes from diploid and allopolyploid species of *Gossypium* (cotton) and from taxa representing its phylogenetic outgroup, *Gossypioides kirkii* and *Kokia kuaiensis* (Seelanan, Schnabel, and Wendel 1997; Wendel et al. 2002). The allopolyploid *Gossypium* species included the commercially important cottons *G. hirsutum* (Upland cotton) or *G. barbadense* (Pima cotton; Sea Island cotton). Allopolyploid cotton contains two, largely colinear (Brubaker, Paterson, and Wendel 1999) genomes (''A'' and ''D'') that were reunited in a common nucleus as a consequence of a remarkable interspecific hybridization event during the Pleistocene (Wendel 1989; Wendel and Cronn 2002), involving two diploids (A genome, D genome) that had evolved in isolation in different hemispheres for perhaps 5–10 Myr (Cronn et al. 2002*a*; Wendel and Cronn 2002). Included in the present study were the closest living models of the diploid progenitors, namely *Gossypium herbaceum* and *Gossypium arboreum* (A genome) and *Gossypium raimondii* (D genome). These two diploids vary nearly twofold in genome size (2C = 2.0 pg and 3.8 pg for the D and A genomes, respectively [see Endrizzi, Turcotte, and Kohel 1985]); these differences are maintained in the derivative allopolyploid (2C = 5.8 pg), which for this and other reasons exhibits near-exclusive bivalent pairing at meiosis (Endrizzi, Turcotte, and Kohel 1985; Wendel and Cronn 2002). The phylogenetic outgroups selected have genomes nearly half as small again as the smallest cotton genome, i.e., 2C = 1.2 pg (Wendel et al. 2002). DNA was isolated from young leaves using published methods (Paterson, Brubaker, and Wendel 1993; Tel-zur 1999) or the Qiagen DNeasy Plant kit following the manufacturer's protocol. In selecting genes to include we focused on those for which robust evidence of orthology could be obtained. The precise nature of this evidence varied among genes but included Southern hybridization against genomic digests (data not shown) to verify single-copy status under high-stringency wash conditions, phylogenetic analysis of sequence data to evaluate the expected concordance of true orthologs with the established organismal history (Wendel and Albert 1992; Wendel 1995; Wendel and Cronn 2002), and comparative mapping to confirm that the genes isolated mapped to equivalent positions in the colinear genomes (Reinisch et al. 1994; Brubaker, Paterson, and Wendel 1999). An additional important criterion was the ability to readily polymerase chain reaction (PCR)-amplify the gene from all species studied. The 28 sets of orthologs selected represent a diversity of genes, including transcription factors, enzymes such as alcohol dehydrogenase and cellulose synthase, and a number of others putatively identified to function based on data-

Key words: molecular evolution, *Gossypium*, C value, DNA content variation.

Address for correspondence and reprints: Jonathan F. Wendel, Department of Botany, Iowa State University, Ames, Iowa 50011. E-mail: jfw@iastate.edu.

Mol. Biol. Evol. 19(12):2346–2352. 2002

base searches (table 1). Some of the genes are described more fully elsewhere (Cronn, Small, and Wendel 1999; Small and Wendel 2000; Cedroni et al. 2002).

Primers for PCR amplification and sequencing were designed as described previously (Cronn, Small, and Wendel 1999; Small and Wendel 2000; Cedroni et al. 2002) or from cotton EST sequences in GenBank. Amplification and sequencing primers are available at J. Wendel's web site (http://www.botany.iastate.edu/~jfw/HomePage/jfwdata_sets.html). In general, two different amplification protocols were used on MJ Research thermocyclers. The first was a ''touchdown PCR'' method: 94°C for 3 min, followed by 10 cycles of 94°C for 1 min, 56°C for 1 min, and 72°C for 2.5 min accompanied by a 0.6°C decrease in annealing temperature each cycle, followed by 25 cycles of 94°C for 1 min, 50°C for 1 min, and 72°C for 2.5 min, ending with a 72°C final extension for 7 min. Other genes were amplified using an initial hot-start of 94°C for 3 min followed by 30 cycles of 94°C for 30 s, 54°C for 30 s, and 72°C for 1 min and 15 s, ending with a final extension at 72°C for 6 min. Annealing temperatures ranged among genes, however, from 48°C to 66°C, and hence, the general amplification conditions given above were adjusted on a gene-by-gene basis when necessary. Sequences that amplified with difficulty were cloned using standard TA cloning protocols and then sequenced from plasmid vectors. Automated sequencing was conducted using the ABI Big Dye v. 2.0 fluorescent primers and ABI Prism 377-3700 systems at the Iowa State DNA Sequencing and Synthesis Facility.

For each gene studied, the allopolyploid species contained two homoeologous sequences, representing descendants of those contributed by the A and D genome donors at the time of polyploid formation. To isolate both homoeologs we cloned amplification products and identified the two duplicates by restriction site analysis, or used homoeolog-specific amplification primers, or discovered both copies following screens of bacterial artifical chromosomes (BAC) libraries from *G. hirsutum* cv. Acala Maxxa (Tomkins et al. 2001) and *G. barbadense* cv. Pima S6 (A. Paterson, personal communication). Since each BAC contained only one of the two homoeologs, this latter strategy proved particularly effective against the nagging problem of in vitro PCR recombination (Cronn et al. 2002*b*).

Sequences were aligned using BioEdit v. 5.0.9 (Hall 1999) and analyzed for substitutions using DnaSP v. 3.53 (Rozas and Rozas 1999). Alignment of orthologs was straightforward due to the low levels of sequence divergence among the taxa studied. Substitution rates for orthologous exons in A and D genome species averaged 3.8% and 0.8% for synonymous and nonsynonymous sites, respectively, across the genes studied, with values approximately twice this size in comparisons with the outgroup. Intron divergence was slightly lower than that of synonomous sites in exons, averaging 3% across orthologous introns in A and D genome cottons and twice this amount in comparisons of either diploid with the outgroup. This low level of sequence divergence additionally facilitated inference of orthologous exon-intron boundaries among the genomes studied. Splice sites were inferred primarily through direct comparisons of genomic sequences with the orthologous cotton cDNAs from which the original PCR amplification primers were designed. For some genes, splice sites were inferred from BLAST searches against other EST databases, as described (Cronn, Small, and Wendel 1999; Small and Wendel 2000; Cedroni et al. 2002).

Intron sizes were inferred for partial or full-length genes for 28 sets of orthologs. As shown in table 1, intron number varied widely among the genes analyzed, ranging from 1 (14 genes) to 11 (*CesA1*). Totalled across the 28 genes, 76 introns were both unambiguously inferred and sequenced from the 5 genomes (3 diploid, 2 homoeologous genomes in the allopolyploid), although only 56 of these were obtained from the outgroup. *Gossypium* introns ranged in size from 71 bp (*AdhC*) to more than 918 bp (a partial intron from *A1550*, a putative aldehyde dehydrogenase) with a mean length of 149.5 ± 151.4 and a median length of 94 bases. These estimates compare closely to the mean (Arabidopsis Genome Initiative 2000) and median (Yu et al. 2002) intron size estimates of 168 and 100, respectively, for a near-exhaustive sampling of genes from *Arabidopsis*, but rice introns apparently are larger (mean and median of 356 and 138, respectively, Yu et al. 2002). We note that both of these model organisms have genomes that are much smaller than the *Gossypium* species studied here, yet their mean intron size is larger.

With respect to the primary issue of whether genome and intron sizes are correlated within *Gossypium*, the data of table 1 show unequivocally that these two genomic features are uncoupled. For homologous and complete introns, the difference in cumulative intron length between the A genome (11,357 bp) and D genome (11,368 bp) diploids was only 11 nucleotides, with the smaller genome having the negligibly higher (0.1%) number. Moreover, there was no case among the 76 introns scored where intron sizes differ significantly between the diploid cottons, with all but two introns (numbers 5 and 6 of C4 kinase) differing by 8 bp or less. Similarly, total intron lengths for any given gene did not differ between the genomes studied. These results extend to the polyploid level, where the data show that intron sizes for homoeologous genes in allopolyploid cotton do not differ appreciably from each other or from those of their diploid progenitors. This latter finding is novel, though not unexpected given earlier, related results (Cronn, Small, and Wendel 1999; Small and Wendel 2000). When data are tabulated for the subset of 56 homologous introns sequenced in either outgroup genus *Gossypioides* or *Kokia*, both of which have much smaller genomes than *Gossypium*, the same general conclusions are reached, with mean intron sizes in *Gossypium* and its outgroup differing in length by an average of two nucleotides (means of 161.2 and 159.2, respectively). Thus, the rate of indel accumulation in introns was relatively low, with no evident differences among taxa in this respect.

Although we sampled only a tiny fraction of the introns in the *Gossypium* genome, the near-identity of

Table 1
Intron Sizes in *Gossypium* and Its Phylogenetic Outgroup

Gene	cDNA or BLAST Match	GenBank Numbers/ Protein Encoded	Length (bp)	Intron		Intron Size (bp)				
				Number	Position	A	A_t	D	D_t	O
A6	GA_Ea0003M07f	AF517646–AF517650/putative protein	522	1 [a]	362–522	161	160	153	153	148
B5	GA_Ea0008D22f	AY116167–AY116171/putative germin E protein precursor	669	1	48–153	106	104	98	98	91
B8	GA_Ea0010N12f	AY115496–AY115500/putative CAAX prenyl protease	489	1	52–274	222	222	221	222	223
C7	GA_Ea0015F06f	AY117065–AY117069/putative SAH7 protein	839	1	69–548	471	471	478	478	390
D1	GA_Ea0017C01f	AY117110–AY117114/putative LIM-domain transcription factor	355	1 [a]	1–35	35	35	35	35	35
				2	71–185	115	115	115	115	115
				3 [a]	276–355	75	75	79	77	78
				Total		225	225	229	227	228
D2	GA_Ea0017H13f	AY116162–AY116166/putative IAA-responsive protein 9	522	1	161–240	79	79	80	80	79
				2	433–509	73	73	75	74	73
				Total		152	152	155	154	152
D5	GA_Ea0017N07f	AY117070–AY117074/putative transporter protein	909	1	16–97	82	82	82	82	82
D7	GA_Ea0018G05f	AY117080–AY117084/putative root hair defective 3 (RHD3)	864	1	1–457	456	455	423	444	448
E9 [b]	GA_Ea0023A19f	AF521968–AF521972/putative kinase-associated protein phosphatase	394	1 [a]	304–394	85	85	91	91	63
F4	GA_Ea0024M11f	AY117095–AY117099/putative ethylene receptor	366	1	70–163	94	94	94	94	94
F8	GA_Ea0025J02f	AY117100–AY117104/putative sugar transporter	359	1 [a]	1–62	62	62	62	62	62
				2	129–239	111	111	111	111	110
				3 [a]	306–359	52	52	53	53	53
				Total		225	225	226	226	225
G11	GA_Ea0029C08f	AY116152–AY116156/putative sulfate transporter	344	1	9–88	80	80	80	80	80
A1550	BI967949	AF201889–AF201893/putative aldehyde dehydrogenase	1,446	1 [a]	1–927	918	916	916	915	869
A1623	GA_Eb0031A04f	AF139474–AF139478/ARF GAP-like zinc finger	719	1	69–234	162	162	164	164	166
				2	346-422	77	77	77	77	77
				3	540-640	101	101	101	101	101
				Total		340	340	342	342	344
A1834	NP_196902	AF139452–AF139456/putative alpha-mannosidase	882	1 [a]	1–770	768	768	768	766	n/a
AdhA	AI727003 GA_Eb0028N03f	AF085064, AF090146, AF136458, AF136459/alcohol dehydrogenase	951	1	7–85	79	79	78	78	n/a
				2	133–206	74	74	72	72	n/a
				3	619–702	84	84	84	84	n/a
				4	779–875	89	97	90	89	n/a
				Total		326	334	324	323	n/a
AdhB	GA_Ed0013D04r	AF226632–AF226635/alcohol dehydrogenase	1,531	1	110–193	83	83	84	84	n/a
				2	241–331	90	90	91	91	n/a
				3	658–778	118	115	115	108	n/a
				4	862–961	100	100	100	100	n/a
				5	1038–1146	109	109	107	107	n/a
				6	1209–1351	143	143	143	143	n/a
				7 [a]	1448–1531	84	84	84	84	n/a
				Total		727	724	724	717	n/a

Table 1
Continued

Gene	cDNA or BLAST Match	GenBank Numbers/ Protein Encoded	Length (bp)	Intron		Intron Size (bp)				
				Number	Position	A	A_t	D	D_t	O
AdhC [c]	GA_Ed0013D04r	AF036568–AF036569, AF036574, AF036575/alcohol dehydrogenase	1,679	1	109–273	149	149	156	165	n/a
				2	321–508	179	179	179	179	n/a
				3	835–920	84	85	81	81	n/a
				4 [a]	1004–1063	60	60	59	59	n/a
				5 [a]	1247–1266	20	20	20	20	n/a
				6	1329–1513	177	177	183	175	n/a
				Total		669	670	681	679	n/a
AdhD	AW508249	AF059418, AF250203–AF250205/alcohol dehydrogenase	1,556	1	9–111	103	104	105	105	n/a
				2	159–427	256	259	259	258	n/a
				3	754–844	89	91	93	93	n/a
				4	916–1206	278	280	280	281	n/a
				5	1283–1373	91	93	92	93	n/a
				6	1436–1522	87	89	89	89	n/a
				Total		904	916	918	919	n/a
GhCLK1 [d]....	BM952576	AY124072–AY124076/protein-associated kinase	2,293	1	52–176	124	124	125	125	125
				2	262–575	312	312	312	312	311
				3	708–790	83	83	83	83	83
				4	905–1046	140	140	140	140	141
				5	1086–1308	221	221	212	212	160
				6	1394–2108	647	648?	670	645?	713
				7	2163–2257	91	91	95	95	93
				Total		1,618	1,619	1,637	1,612	1,626
CesA1	U58283	AF139442–AF139445, AF201886/cellulose synthase A1	4,025	1	140–312	172	172	173	173	173
				2	387–474	86	86	86	86	88
				3	527–620	92	93	93	94	94
				4	766–903	130	130	130	130	138
				5	1171–1283	113	113	113	113	113
				6	1630–1717	88	88	88	88	88
				7	1982–2072	88	88	88	88	91
				8	2286–2379	94	94	94	94	94
				9	2642–2732	91	91	91	91	91
				10	2933–3044	112	112	112	112	112
				11	3399–3493	93	93	86	86	93
				Total		1,159	1,160	1,154	1,155	1,175
CesA2 [b].....	U58284	AF139447–AF139450, AF201887/cellulose synthase A2	2,234	1	125–301	133	133	133	133	177
				2	515–599	85	84	84	84	85
				3	1119–1251	133	133	133	133	133
				4	1603–1716	114	112	111	111	111
				Total		465	462	461	461	506
G1121.....	AF377872	AF139432–AF139435/G1121 protein	749	1	96–435	342	342	342	342	312
GhMYB1.....	GA_Eb0006B13	AY115501–AY115505/MYB-like transcription factor	1,006	1	134–222	86	86	86	86	89
				2	353–444	89	89	92	92	91
				Total		175	175	178	178	180
GhMYB2.....	GA_Ea0026A23	AY115506–AY115510/MYB-like transcription factor	551	1	100–177	78	78	77	77	77

Table 1
Continued

Gene	cDNA or BLAST Match	GenBank Numbers/ Protein Encoded	Length (bp)	Intron		Intron Size (bp)				
				Number	Position	A	A_t	D	D_t	O
GhMYB3.....	GA_Ea0004J09	AF377307, AF377308, AF377316, AF377318, AY115511/MYB-like transcription factor	662	1	134–213	80	80	80	80	80
GhMYB5.....	AF377318	AY115512–AY115516/MYB-like transcription factor	1,057	1	161–260	100	100	100	100	95
				2	391–486	82	82	88	89	96
				Total		182	182	188	189	191
GhMYB6.....	AF034134	AY115517–AY115521/MYB-like transcription factor	1,032	1	141–238	94	94	93	97	98
				2	369–526	153	153	155	155	130
				Total		247	247	248	252	228
				Overall		11,357	11,368	11,368	11,356	

Note.—A total of 76 introns from 28 genes were analyzed. Total genomic length analyzed and intron positions in our alignments are shown for each gene. Homologous sequences were determined from five genomes: A and D genome diploids ("A", "D"), as exemplified by *G. arboreum* or *G. herbaceum* (A) and *G. raimondii* (D); the two homoeologous genomes in allopolyploid cotton ("A_t", "D_t"), represented by *G. hirsutum* or *G. barbadense*; and an outgroup species, *G. kirkii* or *K. kuaiensis* ("O"). n/a = not available. Overall totals at the bottom are shown for *Gossypium* introns only, due to incomplete sampling of homologous introns from the outgroup.

[a] Denotes a partial intron.

[b] *E9* and *CesA2* exhibit hallmarks of pseudogenes.

[c] *AdhC* is a pseudogene in *G. arboreum* due to a large deletion of an exon and parts of two introns. We excluded this deleted region from our analysis, and thus, only portions of introns 4 and 5 are included.

[d] Both the A- and the D genomes harbor a deletion that encompasses the "GT" intron splice start signal. The resulting uncertainty regarding intron length is denoted by a question mark.

intron sizes across taxa varying twofold in genome size and the uniformity of this observation across genes suggests that our primary conclusion is robust, i.e., that intron and genome size evolution are uncoupled in *Gossypium*. It well may be that this will turn out to be common in plants, noting again the comparison of intron sizes in *Arabidopsis*, *Oryza*, and *Gossypium*. Most researchers in animals have focused on broader evolutionary scales than that encompassed here, with the notable exception of Moriyama, Petrov, and Hartl (1998), who compared the sizes of 115 orthologous introns in two *Drosophila* species that vary twofold in genome size, much as in the present study. They reported that *D. virilis*, with a genome size of 0.34 to 0.38 pg, had introns significantly larger (mean of 394 bp) than those of *D. melanogaster* (mean of 283 bp), which has the smaller genome (0.18–0.21 pg). Additional studies are needed to evaluate the generality of this difference between insects and plants with respect to intron and genome size correlation.

One explanation for intron size differences among organisms is that they vary with respect to inherent mutational processes that generate insertions and deletions (Ogata, Fujibuchi, and Kanehisa 1996; Moriyama, Petrov, and Hartl 1998; Petrov et al. 2000; Petrov 2001). In the present study, either divergence amounts were too low to detect subtle differences in deletional bias or such differences do not exist in the lineages examined. In humans, shorter introns have been shown to have more of a mutational bias toward deletions than do longer introns (Vinogradov 2002), suggesting a causal connection between intron size and relative rates of indel accumulation. Carvalho and Clark (1999), in noting that the strength of natural selection should be related to recombination rate, showed a biased occurrence of longer introns in *D. melanogaster* in regions of low recombination, consistent with the notion that larger introns are slightly deleterious. Comeron and Kreitman (2000), however, propose that insertions that create longer introns are selectively advantageous in regions of low recombination precisely because they enhance recombination, thereby counterbalancing the mutational bias toward deletions. More recently, it has been suggested that the association between intron size and recombination rate is a passive response to differences in effective population size, without having to invoke natural selection at the level of the gene (Lynch 2002). The studies cited underscore the complexity of the issue, with determinants of intron size reflecting a balance of evolutionary forces potentially operating at the population, whole genome (Petrov 2001), and genic levels.

It was noted earlier that the correlations between genome and intron sizes that exist at the broader phylogenetic scale (e.g., human vs. avian) are relatively weak (Deutsch and Long 1999) and that "other factors are likely to be involved in the evolution of intron size" (loc. cit., page 3226). Moreover, whether a correlation is observed clearly depends on the taxa studied as well as the phylogenetic scale; maize and humans, for example, have rather similar genome sizes, but introns in humans are on average an order of magnitude larger

than those in maize. As noted by others (Wong et al. 2000; Yu et al. 2002), this difference in gene organization reflects one of the most obvious differences between plant and mammalian genomes, with most transposable element insertions occurring between genes in the former (SanMiguel et al. 1996; Bennetzen 1998; Bennetzen 2000) but within genes (introns) in the latter (Wong, Passey, and Yu 2001). Thus, differences in TE activity and insertional preference likely explain much of the observed correlation between genome and intron sizes in the broader phylogenetic surveys (e.g., Hughes and Hughes 1995; Deutsch and Long 1999; Vinogradov 1999; McLysaght et al. 2000).

For comparisons among more narrowly circumscribed groups, the proximate and ultimate causes of intron size evolution are likely to be more subtle and may reflect the balance of several or more underlying mechanisms as well as external and internal evolutionary forces (Petrov 2001). Moriyama, Petrov, and Hartl (1998) interpreted the longer introns in *D. virilis* compared with the introns in *D. melanogaster* to suggest that mechanisms governing genome size change ''operate more or less uniformly'' throughout the genome. The present study demonstrates that this need not be the case; intron sizes in plants may remain remarkably static even when confronted with mechanisms that massively expand (or contract—Wendel et al. 2002) other genomic components. An important corollary, with general significance to the issue of C-value evolution, is that genome size expansion and contraction likely reflect heterogeneous forces and mechanisms that need not uniformly affect noncoding genomic constituents.

Acknowledgments

We are grateful to R. Noyes, A. Paterson, and J. Rong for picking BAC clones, T. Wilkins and K. Shockey for *GhCLK1* sequences, C. Grover for *CesA* sequence data, R. Percifield for technical assistance, and K. Adams for discussion. Financial support was provided by the National Science Foundation, the US-Israel Binational Science Foundation, the Plant Science Institute of Iowa State University, and the Spanish Ministry of Education, Culture, and Sports.

LITERATURE CITED

ARABIDOPSIS GENOME INITIATIVE. 2000. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature **408**:796–815.

BENNETT, M. D., A. V. COX, and I. J. LEITCH. 1997. Angiosperm DNA c-values database. http://www.rbgkew.org.uk/cval/database1.html.

BENNETZEN, J. L. 1998. The structure and evolution of angiosperm nuclear genomes. Curr. Opin. Plant Biol. **1**:103–108.

———. 2000. Transposable element contributions to plant gene and genome evolution. Plant Mol. Biol. **42**:251–269.

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.

CARVALHO, A. B., and A. G. CLARK. 1999. Intron size and natural selection. Nature **401**:344.

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. (in press).

COMERON, J. M., and M. KREITMAN. 2000. The correlation between intron length and recombination in *Drosophila*. Dynamic equilibrium between mutational and selective forces. Genetics **156**:1175–1190.

CRONN, R. C., M. CEDRONI, T. HASELKORN, C. GROVER, and J. F. WENDEL. 2002*b*. 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*a*. Rapid diversification of the cotton genus (*Gossypium*: Malvaceae) revealed by analysis of sixteen nuclear and chloroplast genes. Am. J. Bot. **84**:707–725.

CRONN, R., R. L. SMALL, and J. F. WENDEL. 1999. Duplicated genes evolve independently following polyploid formation in cotton. Proc. Natl. Acad. Sci. USA **96**:14406–14411.

DEUTSCH, M., and M. LONG. 1999. Intron-exon structure of eukaryotic model organisms. Nucleic Acids Res. **27**:3219–3228.

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

GOFF, S. A., D. RICKE, T.-H. LAN et al. (52 co-authors). 2002. A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*). Science **296**:92–100.

GREGORY, T. R. 2001. Animal genome size database . http://www.genomesize.com.

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.

HUGHES, A. L., and M. K. HUGHES. 1995. Small genomes for better flyers. Nature **377**:391.

LYNCH, M. 2002. Intron evolution as a population-genetic process. Proc. Natl. Acad. Sci. USA **99**:6118–6123.

MCLYSAGHT, A., A. J. ENRIGHT, L. SKRABANEK, and K. H. WOLFE. 2000. Estimation of synteny conservation and genome compaction between pufferfish (*Fugu*) and human. Yeast **17**:22–36.

MORIYAMA, E. N., D. A. PETROV, and D. L. HARTL. 1998. Genome size and intron size in *Drosophila*. Mol. Biol. Evol. **15**:770–773.

OGATA, H., W. FUJIBUCHI, and M. KANEHISA. 1996. The size differences among mammalian introns are due to the accumulation of small deletions. FEBS Lett. **390**:99–103.

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 Mol. Biol. Rep. **11**:122–127.

PETROV, D. A. 2001. Evolution of genome size: new approaches to an old problem. Trends Genet. **17**:23–28.

PETROV, D. A., T. A. SANGSTER, J. S. JOHNSTON, D. L. HARTL, and K. L. SHAW. 2000. Evidence for DNA loss as a determinant of genome size. Science **287**:1060–1062.

REINISCH, A. J., J. DONG, C. L. BRUBAKER, D. M. STELLY, J. F. WENDEL, and A. H. PATERSON. 1994. A detailed RFLP map of cotton, *Gossypium hirsutum* × *G. barbadense*: chromosome organization and evolution in a disomic polyploid genome. Genetics **138**:829–847.

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

SANMIGUEL, P., A. TIKHONOV, Y. K. JIN et al. (11 co-authors). 1996. Nested retrotransposons in the intergenic regions of the maize genome. Science **274**:765–768.

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

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.

TEL-ZUR, N. 1999. Modified CTAB procedure for DNA isolation from epiphytic cacti of the genera *Hylocereus* and *Selenicerus* (Cactaceae). Plant Mol. Biol. Rep. **17**:249–254.

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.

VINOGRADOV, A. E. 1999. Intron-genome size relationship on a large evolutionary scale. J. Mol. Evol. **49**:376–384.

———. 2002. Growth and decline of introns. Trends Genet. **18**:232–236.

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

———. 1995. Cotton. Pp. 358–366 *in* N. SIMMONDS and J. SMARTT, eds. Evolution of crop plants. Longman, London.

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. Syst. Bot. **17**:115–143.

WENDEL, J. F., and R. C. CRONN. 2002. Polyploidy and the evolutionary history of cotton. Adv. Agron. (in press).

WENDEL, J. F., R. C. CRONN, J. S. JOHNSTON, and H. J. PRICE. 2002. Feast and famine in plant genomes. Genetica **115**:37–47.

WONG, G. K.-S., D. A. PASSEY, Y.-Z. HUANG, Z. YANG, and J. YU. 2000. Is ''junk'' DNA mostly intron DNA? Genome Res. **10**:1672–1678.

WONG, G. K.-S., D. A. PASSEY, and J. YU. 2001. Most of the human genome is transcribed. Genome Res. **11**:1975–1977.

YU, J., S. HU, J. WANG et al. (97 co-authors). 2002. A draft sequence of the rice genome (*Oryza sativa* L. ssp. *indica*). Science **296**:79–91.

KENNETH WOLFE, reviewing editor

Accepted 27 August, 2002