Cloning and characterization of the majority of repetitive DNA in cotton (*Gossypium* L.)

Xinping Zhao, Rod A. Wing, and Andrew H. Paterson

Abstract: Repetitive DNA elements representing 60–70% of the total repetitive DNA in tetraploid cotton (*Gossypium barbadense* L.) and comprising 30–36% of the tetraploid cotton genome were isolated from a genomic library of DNA digested with a mixture of four blunt-end cutting restriction enzymes. A total of 313 clones putatively containing nuclear repetitive sequences were classified into 103 families, based on cross hybridization and Southern blot analysis. The 103 families were characterized in terms of genome organization, methylation pattern, abundance, and DNA variation. As in many other eukaryotic genomes, interspersed repetitive elements are the most abundant class of repetitive DNA in the cotton genome. Paucity of tandem repeat families with high copy numbers ($>10^4$) may be a unique feature of the cotton genome as compared with other higher plant genomes. Interspersed repeats tend to be methylated, while tandem repeats seem to be largely unmethylated in the cotton genome. Minimal variation in repertoire and overall copy number of repetitive DNA elements among different tetraploid cotton species is consistent with the hypothesis of a relatively recent origin of tetraploid cottons.

Key words: genome analysis, genome evolution, tandemly repetitive DNA sequences, interspersed repetitive DNA sequences, polyploid.

Résumé : Des éléments d'ADN répété représentant 60–70% de tout l'ADN répété et constituant 30–36% du génome chez le cotonnier tétraploïde (*Gossypium barbadense* L.) ont été isolés d'une banque génomique préparée suite à la digestion de l'ADN avec un mélange de quatre enzymes produisant des bouts francs. Un total de 313 clones contenant vraisemblablement des séquences d'ADN répété d'origine nucléaire ont été classifiés en 103 familles suite à des analyses d'hybridations croisées et Southern. Ces 103 familles ont été caractérisées en fonction de leur organisation génomique, de leur méthylation, de leur abondance et de leur variation au niveau de la séquence nucléotidique. Comme dans plusieurs autres génomes eucaryotiques, les séquences répétées dispersées forment la classe la plus abondante chez le cotonnier. Par contre, ce génome se distingue des autres génomes végétaux par le faible nombre de familles d'ADN répété en tandem en copies multiples ($>10^4$). Dans le génome du cotonnier, les séquences répétées dispersées tendent à être méthylées alors que celles répétées en tandem sont largement non-méthylées. La variation minime quant au répertoire et quant au nombre de copies global des éléments d'ADN répété entre les différentes espèces tétraploïdes du cotonnier sont en accord avec l'hypothèse selon laquelle les cotonniers tétraploïdes seraient d'origine relativement récente.

Mots clés : analyse génomique, évolution des génomes, séquences d'ADN répétées en tandem, séquences d'ADN répétées dispersées, polyploïde.

[Traduit par la Rédaction]

Introduction

Most higher plant genomes are comprised largely of repetitive DNA sequences (Bennett and Smith 1976; Murray et al. 1978, 1979; Flavell 1980, 1982*a*, 1982*b*; Thompson and Murray 1981). Nuclear DNA content can vary widely among plant species, even within the same family, and much of this variation in nuclear DNA content seems to be due to variation in the amount of repetitive DNA (Flavell et al. 1974; Murray et al. 1978, 1979; Flavell 1982*a*, 1982*b*). Therefore, cloning and characterization of repetitive

Corresponding Editor: A.J. Hilliker.
Received April 12, 1995. Accepted July 28, 1995.

X. Zhao,[1] **R.A. Wing, and A.H. Paterson.**[2] Department of Soil and Crop Sciences, Texas A&M University, College Station, TX 77843-2474, U.S.A.

[1] Present address: University of Michigan Medical Center, C568 MSRBII, Ann Arbor, MI 48109-0672, U.S.A.

[2] Author to whom all correspondence should be addressed (tel: (409)845-3773; fax: (409)845-0456; e-mail: ahp2343@bioch.tamu.edu).

sequences is an efficient means of studying the majority of a eukaryotic genome.

While the function of repetitive DNA still remains a subject for speculation (Lewin 1982; Hastie 1985), repetitive sequences are very useful in both evolutionary studies and genome mapping. Specifically, repetitive DNA sequences have been used to infer phylogenetic relationships among related taxa (Miller et al. 1988; Zhao et al. 1989; Dvořák and Zhang 1990; Fabijanski et al. 1990; Crowhurst and Gardner 1991; Zhao and Kochert 1993*b*), and for genetic mapping (Levy et al. 1991; Siracusa et al. 1991; Zhao and Kochert 1992, 1993*a*), DNA fingerprinting (Jeffreys et al. 1985; Nakamura et al. 1987; Siracusa et al. 1991; Zhao and Kochert 1993*b*), physical mapping (Nelson et al. 1989; Ledbetter et al. 1990*a*, 1990*b*; Stallings et al. 1990), and gene cloning (Brilliant et al. 1991; Erlich et al. 1991).

The cotton genus, *Gossypium* L., has long been a focus of genetic, systematic, and breeding research. Cotton is the world's leading fiber crop, and its second most important oilseed. *Gossypium* is comprised of about 50 diploid and tetraploid species indigenous to Africa, Central and South America, Asia, Australia, the Galapagos, and Hawaii (Fryxel 1979, 1992). Diploid species are all $n = 13$ and fall into seven different "genome types," designated A through G based on chromosome pairing relationships (Beasley 1942; Endrizzi et al. 1984, 1985). A total of 5 tetraploid ($n = 2x = 26$) species are recognized. All tetraploid cottons exhibit disomic chromosome pairing (Kimber 1961). Chromosome pairing in interspecific crosses between diploid and tetraploid cottons suggests that tetraploids contain two distinct genomes, which resemble the extant A genome of *G. herbaceum* ($n = 13$) and the D genome of *G. raimondii* Ulbrich ($n = 13$), respectively. The world's cotton fiber is produced from four species, *G. arboreum* L. ($n = 13$, A genome), *G. herbaceum* L. ($n = 13$, A genome), *G. barbadense* L. ($n = 26$, AD genome), and *G. hirsutum* L. ($n = 26$, AD genome).

Prior evidence from DNA reassociation kinetics (Walbot and Dure 1976; Wilson et al. 1976; Geever et al. 1989) suggests that 50–65% of the tetraploid cotton genome is composed of low-copy DNA and 35–50% is repetitive DNA. Evidence of interspersed elements, some of which putatively resulted from ancient retrotransposition, is also well documented (Vanderwiel et al. 1993).

As part of a long-term program to understand the organization and evolution of the cotton genome, we describe the cloning and characterization of major repetitive DNA elements from tetraploid cotton. This work complements our recent analysis of the organization of low-copy elements and the characterization of homoeologous relationships and subgenomic origins of the chromosomes of tetraploid cotton (Reinisch et al. 1994). In the present work, 103 families of repetitive DNA elements were characterized for their organization, abundance, and DNA variation among different cotton species.

Materials and methods

Plant material and DNA isolation

Plant materials used in this study (Table 1) include representatives of each wild tetraploid species, as well as several cultivars from each of the two cultivated tetraploid species *G. barbadense* L. and *G. hirsutum* L. Leaf tissue was harvested for DNA isolation as described previously (Paterson et al. 1993).

Genomic library construction and screening

A genomic library was constructed from *G. barbadense* L. var. Pima S6, in pBluescript KS (+) (Stratagene, U.S.A.). To minimize bias in representation of the genome, genomic DNA was partially digested with a mixture of four blunt-end cutting enzymes (*Alu*I, *Dra*I, *Hae*III, and *Hin*cII). DNA fragments of 0.2 to 1 kilobases (kb) were isolated (gel purified) and ligated into the *Eco*RV site of pBluescript KS (+). The ligation mixture was used to transform *Escherichia coli* strain XL1-blue (Stratagene, U.S.A.). Recombinants were selected on Luria broth (LB) plates with ampicillin, X-Gal, and isopropyl-β-D-thiogalactopyranoside and transferred to new LB plates containing ampicillin. A total of 1900 recombinant colonies were transferred to new LB plates by hand with toothpicks. Duplicate filters were made and colony hybridization was carried out to screen repetitive DNA clones using [^{32}P]dCTP labeled Pima S6 genomic DNA prepared by random primer labeling (Feinberg and Vogelstein 1983). Clones that strongly hybridized to Pima S6 genomic DNA were PCR amplified and the products (2.5 μL) were electrophoresed thru 1% agarose gels and blotted onto Hybond N$^+$ membranes (Amersham, U.S.A.) as described elsewhere (Chittenden et al. 1994). Membranes were probed with [^{32}P]dCTP labeled tobacco chloroplast DNA, exposed to X-ray films, stripped with 0.1 M NaOH, and reprobed with [^{32}P]dCTP-labeled Pima S6 genomic DNA. Clones that showed strong hybridization to genomic DNA, but not to chloroplast DNA, were subjected to further analysis.

Southern blotting and hybridization

Genomic DNA was digested with various restriction enzymes (indicated in figure legends) following the supplier's instructions. Electrophoresis and blotting of genomic digests were as described elsewhere (Chittenden et al. 1994). Genomic blots used 5 μg of DNA from *G. barbadense* L. var. Pima S6, digested with each of 13 or 18 restriction enzymes.

To evaluate the organization of individual repetitive DNA probes in the cotton genome, three restriction enzymes (*Eco*RI, *Alu*I, and *Dra*I) were used to digest DNA samples (5 μg) from different species and cultivars (Table 1).

To classify repetitive DNA clones into families, plasmid DNA from putative nuclear repetitive DNA clones was digested with restriction enzymes to release inserts from the recombinant plasmids, electrophoresed on 1% agarose gels, and Southern blotted (Chittenden et al. 1994). Several sets of filters containing nuclear repetitive clones were made, and each set was hybridized to one repetitive DNA sequence at a time. All clones that hybridized with the same clone were considered to belong to a family, and individual families were further verified by comparing Southern blot hybridization patterns on the replica blots containing Pima S6 DNA digested with either 13 or 18 restriction enzymes.

Probe labeling and hybridization of labeled probes to nylon filters were as described elsewhere (Chittenden et al.

Table 1. Cotton species and their genomes.

Species	Genome	Accession	Geographic origin
G. hirsutum	$(AD)_1$	TM 1	U.S.A.
		Coker 315	U.S.A.
		Stoneville 825	U.S.A.
		TAMCOT 2111	U.S.A.
		TAMCOT CAMD-E	U.S.A.
		Paymaster 145	U.S.A.
		Acala 1517-75	U.S.A.
		TX 25	Mexico
		palmeri	Mexico
G. barbadense	$(AD)_2$	AZK 263	U.S.A.
		3-79	U.S.A.
		Pima S6	U.S.A.
		K101	Bolivia
G. darwinii	$(AD)_5$	PW 44	Galapagos Islands
G. tomentosum	$(AD)_3$	GH 3574	Hawaii (U.S.A.)
G. mustelinum	$(AD)_4$	AD 4-7	Brazil

Note: Genome designations follow Endrizzi et al. (1985).

1994). The membranes were washed at 65°C in 1× SSC (0.15 M NaCl plus 0.015 M sodium citrate) and 0.1% SDS for 20 min, 0.5× SSC and 0.1% SDS for 20 min, and 0.5× SSC and 0.1% SDS for 60–90 min. Autoradiography used Fuji X-ray film and a single Spectra L-Plus intensifying screen at −80°C.

Slot blot hybridization

Quantitative slot blot hybridization (Rivin et al. 1986) was used to estimate copy numbers for selected DNA sequence families following Zhao et al. (1989). Nuclear DNA from *G. hirsutum* L. race "palmeri" was used in this experiment. Its genome size was estimated as 4.655 pg/2C (e.g., 2.246×10^9 base pairs (bp) / haploid genome), which is the average of two cultivars of *G. hirsutum* L. (Arumuganathan and Earle 1992).

DNA sequence analysis

Cloned repetitive elements were subjected to cycle sequencing on both strands using either T7 and T3 primers or M13 forward and reverse primers, making use of an automated DNA sequencer (model 373A, version 1.2.0, Applied Biosystems, Foster City, Calif.).

Results

The cotton DNA library was characterized by probing with genomic DNA and chloroplast DNA successively. A total of 1900 recombinants were screened, having an average insert size of 500 bp and representing a total of about 1 million base pairs, or about 0.04% of the tetraploid genome. Three hundred and thirteen nuclear repetitive DNA clones were selected for further study from the library by colony hybridization and Southern blot hybridization on filters containing PCR-amplified recombinants, successively. Based on screening of 0.04% of the genome, it was anticipated that the majority of elements that occur in 2500 copies or more should be represented among these clones.

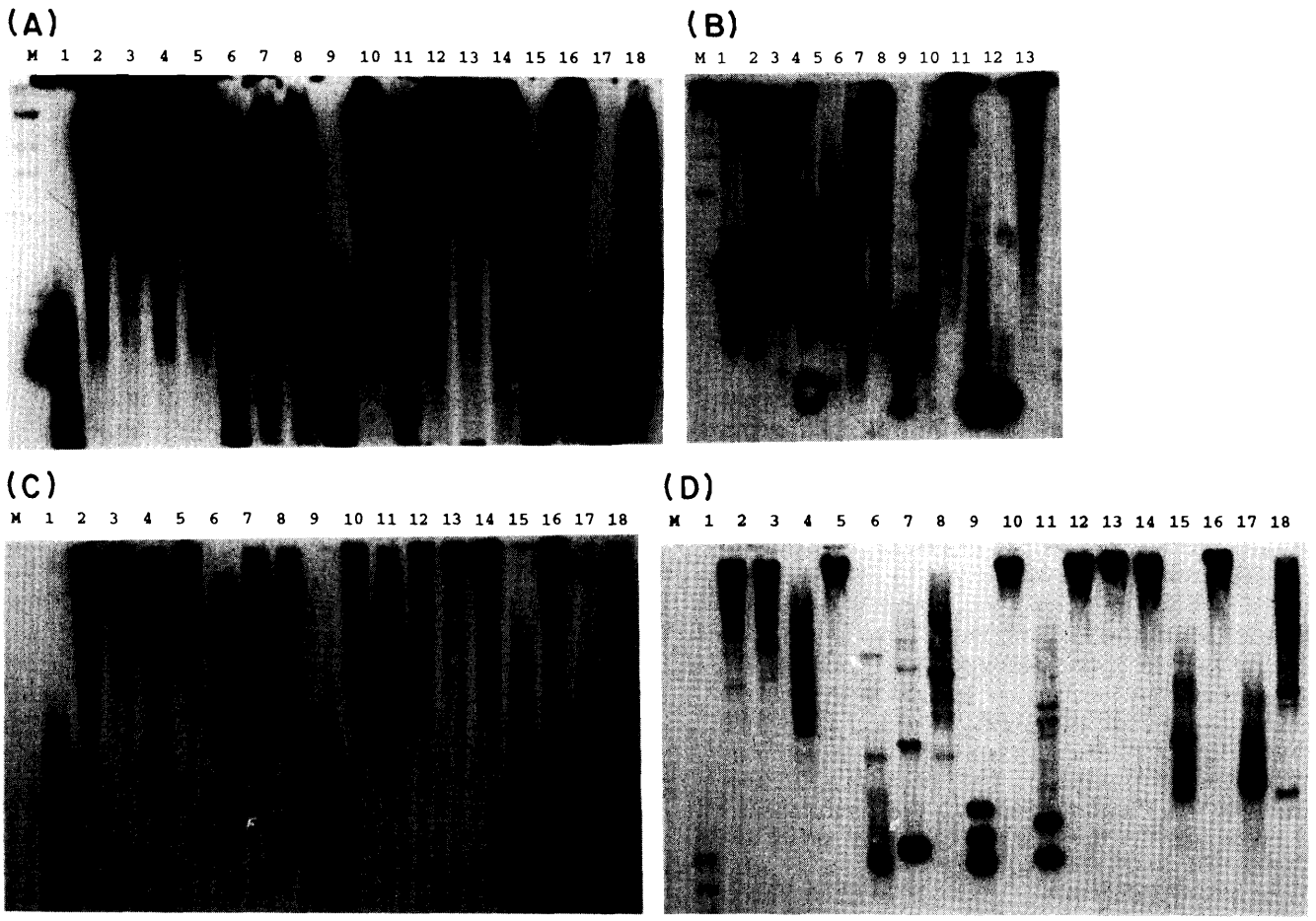
Based on cross hybridization and comparison of Southern hybridization patterns, the 313 repetitive DNA clones were classified into 103 different families of nonoverlapping elements.

Organization of repetitive sequences

Southern blot hybridization on blots containing Pima S6 genomic DNA digested with 13 or 18 restriction enzymes was used to investigate the organization of repetitive DNA in tetraploid cotton. Hybridization patterns were analyzed for each repetitive DNA clone. Figure 1A shows an example of Southern blot hybridization using an interspersed repeat clone, pXP004, as a probe. Individual bands were barely discernible with strong background hybridization on most restriction enzyme digests. In contrast, tandem repeats show one or a few bands with most restriction digests (Fig. 1B). In some instances, tandem repeats were confirmed by a ladder with complete digests (Fig. 1B), indicating point mutations that eliminated the relevant restriction site(s). Tandem organization was further clarified by partial digestion of genomic DNA and confirmation that a ladder of bands was observed (not shown).

Assignment of individual elements to a "tandem" or "interspersed" motif was not always absolute. Some repeat families showed multiple bands on most of the enzyme digests with less background (Fig. 1C) than classical interspersed repeats (Fig. 1A). In addition, for some enzyme digests only a few bands were found (Fig. 1C). These results suggest that some copies of such families are interspersed in the genome, with other copies clustered in a few regions of the genome. We classify such repetitive DNA sequences as "partially interspersed" (PI). A few repetitive sequences showed several hybridization bands with most enzyme digests, with little or no smear hybridization

Fig. 1. Southern blot hybridization analysis showing the organization of repetitive DNA in the cotton genome. Genomic DNA of *G. barbadense* L. var. Pima S6 was digested with (A, C, and D) *Alu*I, *Bam*HI, *Bgl*I, *Bst*NI, *Cfo*I, *Dra*I, *Eco*RI, *Eco*RV, *Hae*III, *Hha*I, *Hin*dIII, *Hpa*II, *Pst*I, *Pvu*II, *Sau*3AI, *Sfi*I, *Taq*I, and *Xba*I (lanes 1–18, respectively) or with (B) *Hpa*II, *Msp*I, *Bam*HI, *Cfo*I, *Dra*I, *Eco*RI, *Eco*RV, *Hae*III, *Hin*dIII, *Pst*I, *Sau*3AI, *Taq*I, and *Xba*I (lanes 1–13, respectively). Digested DNA and molecular weight marker (M, λDNA digested with *Hin*dIII, bands from top to bottom are 23.1, 9.4, 6.6, 4.2, 2.3, 2.0, and 0.6 kb, respectively) were fractionated by electrophoresis in 1% agarose gels, Southern blotted, and hybridized to the cloned repetitive DNA sequences. (A) Typical hybridization patterns generated by probing with an interspersed repeat, pXP004. (B) Typical hybridization patterns obtained by probing with a tandem repeat, pXP031. (C) An example of hybridization patterns generated by probing with a partially interspersed repeat, pXP172. (D) An example of hybridization patterns obtained by probing with a partially tandem repeat clone, pXP128.



(Fig. 1D), suggesting that most species of these families are organized in tandem arrays with a few copies interspersed or with repetition of a mutated sequence in the genome. Such repetitive DNA families were classified as "partially tandem" repeats (PT).

Therefore, each of the two broad categories was divided into two subgroups: interspersed or partially interspersed and tandem or partially tandem. Results for each of the 103 families are summarized in Table 2. A total of 83 (81%) of the 103 repetitive DNA families are interspersed or partially interspersed repeats (I, 43 families; PI, 40 families), and the remaining 20 (19%) are tandem or partially tandem repeats (T, 18 families; PT, 3 families).

Methylation of repetitive sequences

Methylation of repetitive sequences in the cotton genome was investigated by comparing hybridization patterns of genomic blots containing *Msp*I digests with those of *Hpa*II digests. Although *Hpa*II and *Msp*I recognize the same nucleotide sequence (5′ CCGG 3′), *Hpa*II is sensitive to methylation of either of the two cytosines in its recognition sequence, while *Msp*I is sensitive only to methylation of the external cytosine. Therefore, differences in hybridization patterns between these two enzyme digests should reflect variation in the degree of methylation in the target sequences.

The methylation pattern was determined for each of the 103 repeat families and summarized in Table 2. Individual families were either classified as methylated or not methylated (N). Out of the 103 families, 33 (32%) produced identical hybridization patterns with these two restriction enzymes, and therefore, are not methylated. As an example, Fig. 2A shows hybridization patterns generated by probing with such a clone (pXP039). However 70 (68%) of the 103 repeat families showed different hybridization patterns

Table 2. Methylation of cotton repetitive DNA sequences.

Organization[a]	Methylation patterns of repeat families[b] H	M	L	N	Total
I	4	11	17	11	43
PI	7	15	13	5	40
T	0	2	0	15	17
PT	0	1	0	2	3
Total	11	29	30	33	103

[a]I, interspersed; PI, partially interspersed; T, tandem; PT, partially tandem.
[b]H, highly methylated; M, moderately methylated; L, low methylated; N, not methylated.

with the two enzymes, indicating methylation. Different degrees of methylation were observed for these families. In highly methylated (H) families, virtually all hybridization signal on the *Hpa*II digest was in the high molecular weight range (ca. >10 kb), suggesting virtually no digestion of the repeat with the enzyme. In contrast, most hybridization signal was in a much lower molecular weight range for the *Msp*I digest (Fig. 2B). The relative hybridization signals in the lower molecular weight range for the *Msp*I digest reduced gradually from moderately methylated (M) (Fig. 2C) to low methylated (L) (Fig. 2D).

Most interspersed or partially interspersed families (67 of 83) in cotton were found to be methylated. In contrast, most tandem or partially tandem repeats (17 of 20) were not methylated. However, exceptions were found in each case (Table 2).

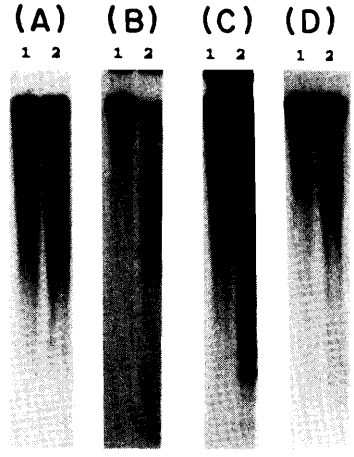
Abundance of repetitive sequences

The relative abundance of repeat families was first estimated to be high, moderate, or low by visual comparison of hybridization signals on Southern blots. All "high-abundance" families (25) and 8 representative "moderate-abundance" families were selected for more rigorous copy number estimation by quantitative slot-blot hybridization. Copy number, and the percentage of the haploid genome comprised by each of the 33 families are shown in Table 3. This determination has an inherent assumption that the cloned sequence is representative of its respective family.

High-abundance families ranged from 15 000 to 100 000 copies, while moderate-abundance families ranged from 4 000 to 10 000 copies. We estimate that sufficient DNA was present on our colony lifts to detect elements occurring in >100 copies, thus low-abundance families are estimated to range from 100 to 4000 copies. Since there is no overlap in copy numbers among the "classes" we determine by visual assessment of hybridization signal on genomic Southern blots (high abundance, moderate abundance, and low abundance), this visual assessment represents a meaningful (albeit approximate) estimation of copy number.

Together, the 33 families for which copy numbers were determined by slot-blot hybridization analysis (see above) constitute 24.5% of the haploid genome of *G. hirsutum* L.

Fig. 2. Southern blot hybridization showing methylation patterns of repetitive DNA in the cotton genome. Genomic blots containing *Hpa*II (lane 1) and *Msp*I (lane 2) digests of *G. barbadense* L. var. Pima S6 were hybridized to the cloned repetitive DNA sequences. Examples of different degrees of methylation are shown: (A) not methylated (N, pXP39), (B) highly methylated (H, pXP016), (C) moderately methylated (M, pXP015), and (D) low methylated (L, pXP027).



Individual families accounted for 0.03–2% of the haploid cotton genome. Further, using copy numbers of the elements analyzed, we can estimate the portion of the genome accounted for by the remaining elements deemed to be of similar abundance by visual assessment of hybridization signals. Forty-six families estimated visually to be moderately abundant would comprise about 7.2% (the mean of the estimated range of 4.0–10.3%, which corresponds to the range in copy number of the moderately abundant class) of the haploid genome using the average copy number for the 6 moderately abundant families (7100) and the average insert size of 500 bp. The 24 low-abundance families would probably have a copy number of less than 10^3 and make up less than 0.5% of the haploid genome. Therefore, the 103 repeat families comprise 29–35% (32.2%) of the haploid genome of *G. hirsutum* L. Hybridization signal from Southern blots of different tetraploid cotton species showed no obvious differences for all the repeat families analyzed, suggesting that this estimate is generally applicable across tetraploid cottons.

Polymorphisms detected by repetitive DNA sequences among cotton species

Variation at individual repetitive DNA sequence loci among five different tetraploid cotton species (Table 1) was analyzed by Southern hybridization. One to two blots of the three restriction enzyme digests (*Eco*RI, *Alu*I, and *Dra*I) were used in hybridization to each of the 103 repetitive sequences. As an example, Fig. 3 shows Southern blot hybridization of *Eco*RI digested genomic DNA of the five tetraploid cotton species to pXP020. Very similar or identical hybridization patterns are obvious among all the five species. Figure 3 also shows some variation in relative abundance of pXP020 in different species. *Gossypium*

Table 3. Characteristics of selected repeat families.

Probe name	Copies per haploid genome[a]	Insert size (bp)	Percent haploid genome[b]	Organization	Methylation
pXP72	100 000	230	1	PI	H
pXP4	100 000	460	2	I	L
pXP20	100 000	380	1.7	PI	L
pXP271	90 000	530	2.1	I	H
pXP224	80 000	280	1	I	M
pXP190	80 000	360	1.3	I	M
pXP117	80 000	150	0.5	PI	H
pXP24	75 000	390	1.3	PI	H
pXP6	70 000	460	1.4	I	M
pXP18	60 000	540	1.4	I	M
pXP137	60 000	480	1.3	PI	M
pXP95	50 000	280	0.6	PI	L
pXP33	50 000	500	1.1	PI	M
pXP167	50 000	330	0.7	T	M
pXP1-54	50 000	460	1	T	N
pXP2-58	40 000	500	0.9	I	L
pXP1-24	40 000	400	0.7	T	N
pXP67	30 000	610	0.8	PI	H
pXP39	20 000	380	0.3	I	N
pXP200	20 000	260	0.2	I	N
pXP2-27	20 000	630	0.6	PT	N
pXP172	20 000	350	0.3	PI	H
pXPx270	15 000	650	0.4	I	L
pXP2-81	15 000	390	0.3	PI	M
pXP2-12	15 000	810	0.5	T	N
pXP2-38	10 000	400	0.2	I	N
pXP128	10 000	760	0.3	PT	M
pXP125	10 000	410	0.2	I	L
pXP1-23	7 000	90	0.03	T	N
pXP31	6 000	400	0.1	T	N
pXP57	5 000	680	0.1	PI	M
pXP2-47	5 000	580	0.1	PI	L
pXP96	4 000	720	0.1	PI	M

[a]Estimated copy numbers of the repeat families.
[b]Percentage of the haploid genome of *G. hirsutum* L. composed of individual families, calculated as copy number × insert size / 2.246×10^9 (Arumuganathan and Earle 1992).

barbadense (lanes 10–13), *G. darwinii* (lane 14), and *G. mustelinum* (lane 16) have fewer copies of the pXP020 repeat than *G. hirsutum* (lanes 1–9). However, no detectable variation in abundance was found on most repetitive families characterized in this study (not shown). Restriction fragment length polymorphisms (RFLPs) among the five tetraploid cotton species were detected on only 18 of the 103 families (17%). These results indicated that there is minimal variation in the repetitive fraction of the genome among various tetraploid cotton species, suggesting a close relationship of these species. The 17 polymorphic repeat families may be useful for distinguishing different tetraploids.

DNA sequence analysis

Figure 4 shows DNA sequences of four repetitive elements, two representing tandem repeats (pXP196, GenBank accession Number U31111; pXP1-3, U31112) and two representing interspersed repeats (pXP1-62, U31113; pXP2-18, U31114). The clone pXP196 contains a 357-bp fragment and has a AT content of 58.5%. There are two classes of imperfect direct repeats present in the cloned sequence, each represented by two elements. One member of each pair overlaps at positions 252–257 (Fig. 4). Although a relatively long open reading frame (ORF) is found in the sequence (positions 223–355), a GenBank search failed to find any sequences showing high homology to the clone at a fragment length of 40 bp or more.

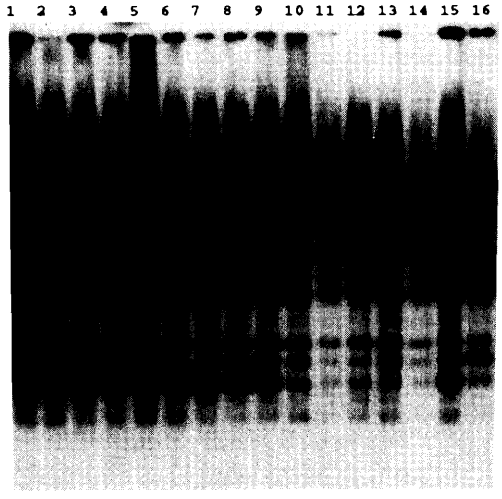
The 275-bp fragment of pXP1-3 is also AT rich (67.6%) and contains three classes of shorter repeats. A striking feature of the sequence is the presence of three perfect repeats (15 bp) located adjacent to each other with one nucleotide between repeat units (positions from 168 to 214, Fig. 4). A second set of imperfect direct repeats is found, with one unit partly overlapping the last of the

three perfect repeats. Finally, the latter unit of a third pair of imperfect repeats (positions 203–217 and 218–232) almost completely overlaps one of the second set of repeats. Two short ORFs were found (positions 18–42 and 98–140) in the sequence but showed no homology to known genes. Two different regions of the sequence show 61% (47 of 77) DNA sequence similarity with the *Sus scrofa* (pig) epsilon-globin gene (GenBank accession Number X86792) and 66% (38 of 57) DNA sequence similarity with the *Bacillus* spp. Na/H transporter (GenBank accession Number D31823), respectively.

The 464-bp fragment of pXP1-62 (63.6% AT) includes two classes of repeats. In one class, a pair of 14-bp inverted repeats (two mismatches) is found (positions 70–97) and two additional units near the end of the element are in the 5′–3′ orientation. The other class is also present as two elements with three mismatches (all C versus G) and starting at positions 255 and 285, respectively. No sequences in the GenBank share homology with pXP1-62 over a length of 60 bp or more.

pXP2-18 contains a DNA fragment of 630 bp (62% AT). Two classes of repeats are found within the element. A pair of direct repeats of 28 bp (with one base mismatch) is separated by one nucleotide, as shown in Fig. 4 (positions 84–111 and 113–140). In addition, four imperfect 12-bp repeats are scattered throughout the sequence. A long ORF is also found (positions 259–630, including two of the 12-bp repeats); however, shared homologies are limited to very short DNA fragments (less than 50 bp), making it difficult to infer functional significance of the sequence.

Fig. 3. Southern blot hybridization analysis showing similar RFLP patterns among various tetraploid cotton species generated by probing with the repetitive DNA sequence. Genomic blot containing *Eco*RI-digested DNA from different cotton species and cultivars was hybridized to the cloned repetitive DNA sequence pXP024. Lanes 1–9: *Gossypium hirsutum* $(AD)_1$: lane 1, TM 1; 2, Coker 315; 3, Stoneville 825; 4, TAMCOT 2111; 5, TAMCOT CAMD-E; 6, Paymaster 145; 7, Acala 1517-75; 8, TX25; 9, palmeri; lanes 10–13: *G. barbadense* $(AD)_2$: 10, AZK 263; 11, 3-79; 12, Pima S6; 13, K101; lane 14: *G. darwinii* $(AD)_5$, PW 44; lane 15: *G. tomentosum* $(AD)_3$, GH 3574; lane 16: *G. mustelinum* $(AD)_4$, AD 4-7. See Table 1.



Discussion

Global cloning and characterization of repetitive DNA in the cotton genome

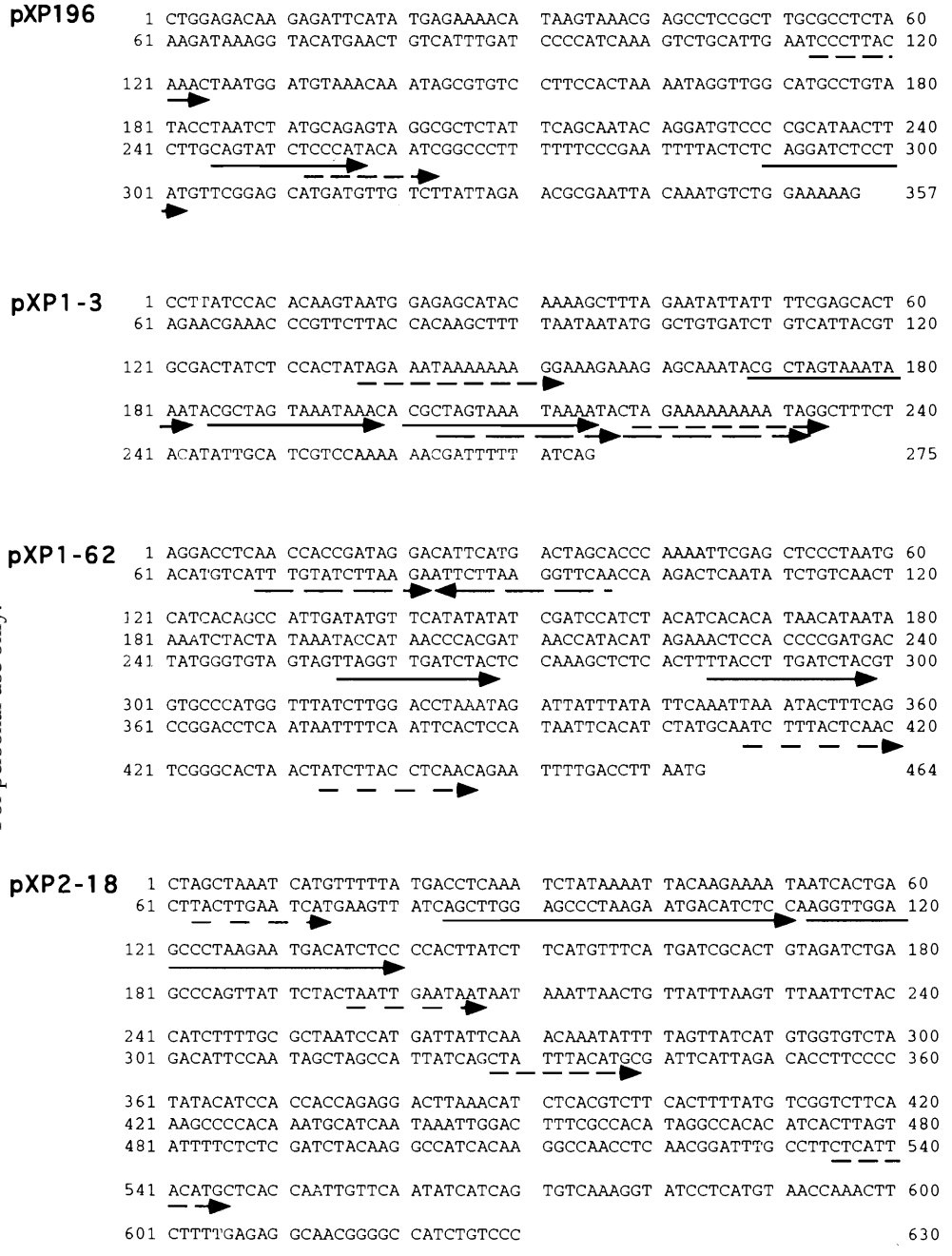
We have isolated 103 families of repetitive DNA elements, which are estimated to comprise 29–35% of the genome, and 60–70% of the repetitive fraction of tetraploid cotton. Therefore, we believe that we have cloned and characterized at least segments of most of the abundant repetitive DNA elements in the tetraploid cotton genome. While cloning of repetitive sequences is routinely done, most previous studies have targeted specific elements digestible with a particular restriction enzyme or have characterized a small portion of the overall repetitive fraction. Our objectives are to understand the overall organization and distribution of repetitive DNA among different cottons and to utilize repetitive DNA as a tool in both global and local analysis of the cotton genome.

To isolate repetitive DNA on a genome-wide scale, a genomic library containing a representative sample of genomic DNA is required. Ganal et al. (1988) isolated and characterized four major repeat families from tomato by constructing and screening a sheared DNA library. Prior analysis showed that the sheared DNA library was a representative sample of the genome and that 22% of the clones in the library contained repetitive DNA (Zamir and Tanksley 1988). The four repeat families characterized accounted for approximately 5% of the total nuclear genome in tomato and 25–50% of all repetitive DNA (Ganal et al. 1988). In this report, we constructed a genomic library from DNA that was partially digested with a mixture of four blunt-end cutting enzymes (*Alu*I, *Dra*I, *Hae*III, and *Hin*cII). All four enzymes have different recognition sites and have a total expected cutting frequency of 1/128 bp. In addition, *Hae*III, *Hin*cII, and probably also *Dra*I (AAATTT), are insensitive to methylation. Therefore, the DNA library should provide a reliable representative sample of the tetraploid cotton genome. Compared with a sheared library, the mixed enzyme partial genomic library is much easier to construct, since it does not require additional molecular manipulation of the DNA fragments such as addition of poly G or of linkers or adapters to the ends of sheared DNA fragments before they can be cloned.

The possibility that different parts of a repeat were cloned into different families (as classified in this study) cannot be ruled out by the available data, but comparison of Southern hybridization patterns of different clones on replica blots containing 13 or 18 enzyme digests suggests that such duplication accounts for few, if any, families. For most repeat families, an insert size of 500 bp fits well with the average size of individual repeat units in most plants (usually 300–800 bp; Flavell 1980; Thompson and Murray 1981; Lapitan 1992).

Repetitive DNA with less than 100 copies would probably not be detected by colony hybridization and is therefore not included in our collection. We know from analysis of ca. 3000 candidate DNA sequences (not shown) for RFLP mapping that numerous low-abundance families do occur

Fig. 4. DNA sequence analysis of four repetitive elements. Direct and inverted repeats are indicated underneath the sequence data. Different subdomains, of related DNA sequence, within each repetitive element, are indicated by solid or dotted lines.



in cotton. In addition, a few more repeat families from other genomic libraries, 5S rDNA and part or complete repeats of 45S rDNA, have been isolated and are not included in this report (unpublished). The figure of 32.2% is therefore a minimum estimate of the repetitive fraction in the tetraploid cotton genome. Independent estimates of the size of the repetitive fraction in tetraploid cotton range from 35 (Walbot and Dune 1976) to 50% (Wilson et al. 1976; Geever et al. 1989). Conservatively, repeat families characterized in this report should account for at least 60% of the repetitive fraction of the tetraploid cotton genome.

General features of the repetitive DNA fraction of cotton

Most repeat families (83 of 103) isolated in this study are interspersed in the cotton genome and have copy numbers that range from 10^3 to 10^4 (except for three families, pXP004, pXP020, and pXP072, with a copy number of 10^5) and thus fall into the moderate-abundance class (10^2–10^5 copies; Bouchard 1982). Therefore, moderately abundant interspersed repetitive DNA comprises the largest class of repetitive sequences in the cotton genome, as in most eukaryotic genomes characterized (Bouchard 1982; Lewin 1982; Hastie 1985).

Tandemly repetitive DNA, the satellite DNA, usually represents the most highly repetitive sequence family in eukaryotic genomes (Singer 1982; Flavell 1982*a*). Tandem repeat families with high copy numbers have been detected and isolated from many plant species (Bouchard 1982; Flavell 1982*a*, 1986; Lapitan 1992; Zhao et al. 1989). Out of the cloned 103 repeat families, only 20 (19%) are tandem repeats. Five were estimated to have a copy number in the range of 10^4 and the rest had a lower copy number. None of the tandem repeats meets the traditional threshold of 10^6 copies (Singer 1982) for classical "highly repetitive" DNA. Satellite DNA tends to have short repeat units and might be more prominent in a sheared library than in our mixed enzyme partially digested library. Alternatively, the cotton genome seems to have few, if any, tandem repeat families with a copy number comparable with the interspersed repeat families having the highest copy number so far identified. This unusual feature of cotton is supported by evidence from DNA–DNA reassociation kinetics, which showed the rapidly reassociating fraction to be small (Wilson et al. 1976; Geever et al. 1989), and buoyant density centrifugation experiments, which showed no detectable satellites (Walbot and Dure 1976). This feature makes the cotton genome different from many other higher plant genomes (Flavell 1982*a*; Lapitan 1992).

While Southern hybridization analysis provides information about the organization of repetitive DNA, the reliability of conclusions drawn from such an analysis depends heavily on the number of restriction enzymes used. A tandem repeat might be considered to be interspersed if the enzymes used do not have recognition sites within the repeat. To minimize this problem, we used a minimum of 13 restriction enzymes in our Southern hybridization analysis. Hybridization patterns on all the restriction enzymes were considered when we analyzed the organization of a given repetitive DNA in the tetraploid genome. This allowed a more detailed comparison of the organization of repetitive sequences to be made. For example, the difference among repeat families within the category of interspersed repeats would permit distinction between typical interspersed and partially interspersed repeats. Therefore, conclusions about the organization of repetitive DNA in the genome are more reliable and better reflect the nature of a repetitive DNA in the genome. In addition, this analysis and comparison among different repeats provide further evidence to clarify the classification of repeat families based on cross-hybridization experiments.

Methylation patterns for all repeat families were analyzed using the isoschizomer pair *Hpa*II and *Msp*I. A total of 70 (68%) of the 103 families are methylated to some degree. Cotton DNA has been estimated to have a G + C + 5 MeC content of 36.1%, of which about 25% of the C (4.6% of the total bases) is the methylated derivative 5 MeC (Ergle and Katterman 1961; Walbot and Dure 1976). Since the majority of repeat families are methylated, methylated C residues may be concentrated in the repetitive fraction of the cotton genome. We note that genomic DNA digestion with *Pst*I, a methylation-sensitive enzyme, is a highly effective means of enriching for 1- to 2-kb digestion products suitable as low-copy DNA probes (Reinisch et al. 1994).

Most families of interspersed repetitive sequences, specifically 67 (81%) of the 83 families, are methylated, while only 3 (15%) families of the 20 tandem repeats are methylated. This biased methylation pattern between the two classes of repetitive DNA may reflect a difference in base composition of DNA or chromosomal location between tandem and interspersed families. The isoschizomer pair, *Hpa*II and *Msp*I, can only detect methylation of C residues. It is possible that some of the tandem repeats are AT enriched in base composition and do not contain the recognition sequence for these two enzymes. Considering the high degree of homology among different repeats of the tandem class, as demonstrated in other cases (Flavell 1980, 1982*a*, 1982*b*; Lapitan 1992), this possibility is very likely for those families that cannot be cut by the two enzymes. However, a single hybridization band with the same molecular weight for the two enzymes was also observed for some tandem repeat families. In such cases, the recognition site for the two enzymes must be within the repeat. It is therefore possible that the C residues existing in such tandem repeat families may be hypomethylated.

Little variation in abundance of repetitive sequences was found among the tetraploid species of *Gossypium*. This suggests that little, if any, net amplification or deletion of individual repetitive sequences in the tetraploid genomes has occurred during the divergence of tetraploid species from a common ancestor 1.1–1.9 million years ago (Wendel 1989). It is well documented that repetitive DNA sequences represent a fraction of the plant genome undergoing rapid changes during the course of evolution (Brutlag 1980; Flavell 1980, 1982*a*, 1982*b*; Singer 1982; Evans et al. 1983; Grellet et al. 1986; Martinez-Zapater et al. 1986; Crowhurst and Gardner 1991). As a result of the rapid changes, species- and genome-specific repetitive DNA sequences are not uncommon and have been isolated from a number of plant species (Bedbrook et al. 1980; Gerlach and Peacock 1980; Rayburn and Gill 1986; Schweizer et al. 1988; Zhao et al. 1989; Hueros et al. 1990; Crowhurst and Gardner 1991; Anamthawat-Jonsson and Heslop-Harrison 1993). The 103 repeat families were estimated to comprise about 29–35% of the tetraploid cotton genome. If polyploid formation had been ancient, discernible variation among these species in abundance for some of the 103 repeat families might have accumulated. Therefore, our results support the conclusion drawn from the study of chloroplast genome variation (Wendel 1989) that tetraploid cottons are of relatively recent origin.

Southern hybridization analysis based on 1–2 enzyme digests showed that most repeat families failed to detect RFLPs among tetraploid species, providing further evidence to support the above conclusion. It is noted that RFLP analysis using repetitive DNA sequences could likely fail to detect any rearrangements of a small number of repeats, considering the repetitive nature of the probes, especially those families with thousands or more copies, since most hybridization bands presumably arise from multiple copies of the repeat. However, such analysis should detect any changes involving the majority of copies of the repeat in the genome, therefore, providing information on genomic relationship between related species.

On the other hand, a few repeat families did detect RFLPs among these tetraploid species. Since little variation in abundance for all repeat families was observed, these RFLPs may result from rearrangements of many copies of the repetitive DNA sequences in related tetraploid genomes.

We note that similar levels of net abundance of repeats across the genome do not necessarily reflect conserved copy number at any single site. Studies to be published separately will describe detailed characterization of locus-specific variation in copy number of simple-sequence repeats ("microsatellites"), which show every indication of being as useful in cotton as in human or other organisms (Beckmann and Soller 1990; Zhao and Kochert 1993*a*).

Sequence analysis of four cotton repetitive elements suggests that many are AT rich, comprised in part of shorter direct and (or) inverted repeats, and contain ORFs. Although representing only a small subset of the total repertoire of repetitive element families in cotton, each of these features is, in principle, consistent with the possibility that replicative transposition may have played a significant role in cotton genome evolution (Vanderwiel et al. 1993). Sequence analysis of the entire repertoire of cloned cotton repetitive element families will afford more detailed characterization of the structure and function of these elements (X. Zhao and A.H. Paterson, in preparation).

The use of repetitive DNA in physical mapping of the cotton genome

Human geneticists have demonstrated that repetitive DNA is very useful for physical mapping (Stallings et al. 1990). The construction of contigs from a given chromosomal region frequently involves the use of repetitive DNA as either hybridization probes or PCR primers for DNA fingerprinting of individual YACs (Nelson et al. 1989; Stallings et al. 1990). Similarly, repeat families reported here provide a rich source of hybridization probes for contig construction in cotton genome mapping. Three interspersed repeat families, pXP004, pXP020, and pXP072, were estimated to have copy numbers of 100 000 (each) in tetraploid cotton, almost as abundant as the *Alu* family in the human genome (Deininger et al. 1981). Based on in situ hybridization (C. Crane, D. Stelly, J. Wendel, X. Zhao, and A.H. Paterson, unpublished data), these three elements are different from each other and from the most prominent retroposon families in cotton (Vanderwiel et al. 1993). Therefore, repeat elements identified in this study provide new tools for physical mapping or painting of cotton chromosomes by in situ hybridization.

The use of repetitive DNA in phylogenetic studies

Although limited DNA variation in the repetitive elements was found among different tetraploid species, our survey of genome specificity showed that these repetitive elements have a wide spectrum of genome distribution. Either RFLPs or polymorphisms in abundance have been detected at different levels among cotton species with A, B, C, D, E, F, G, and AD genomes by the repeat families (X. Zhao and A.H. Paterson, in preparation). Therefore, the cloned repetitive DNA is potentially useful for investigating species and genome relationships in the genus *Gossypium* L.

Acknowledgments

This study was supported by the Texas Higher Education Coordinating Board (999902-148 to A.H.P. and R.A.W.), United States Department of Agriculture, National Research Initiative Competitive Grants Program (91-37300-6570 to A.H.P., 93-37300-8819 to R.A.W. and A.H.P.), and the Texas Agricultural Experiment Station (to A.H.P. and R.A.W.). We thank S. Coomber for providing tobacco chloroplast DNA and D. Altman, E. Percival, R. Percy, and J. Wendel for providing seeds and (or) tissue.

References

Anamthawat-Jonsson, K., and Heslop-Harrison, J.S. 1993. Isolation and characterization of genome-specific DNA sequences in *Triticeae* species. Mol. Gen. Genet. **240**: 151–158.

Arumuganathan, K., and Earle, E.D. 1992. Nuclear DNA content of some important plant species. Plant Mol. Biol. Rep. **9**: 208–218.

Beasley, J.O. 1942. Meiotic chromosome behavior in species hybrids, haploids, and induced polyploids of *Gossypium*. Genetics, **27**: 25–54.

Beckmann, J.S., and Soller, M. 1990. Toward a unified approach to genetic mapping of eukaryotes based on sequence tagged microsatellite sites. Bio/Technology, **8**: 930–932.

Bedbrook, J.R., Jones, J., O'Dell, M., Thompson, R.D., and Flavell, R.B. 1980. A molecular description of telomeric heterochromatin in secale species. Cell, **19**: 545–560.

Bennett, M.D., and Smith, J.B. 1976. Nuclear DNA amounts in angiosperms. Philos. Trans. R. Soc. Lond. Series B, **274**: 227–274,.

Bouchard, R.A. 1982. Moderately repetitive DNA in evolution. Int. Rev. Cytol. **76**: 113–193.

Brilliant, M.H., Gondo, Y., and Eicher, E.M. 1991. Direct molecular identification of the mouse pink-eyed unstable mutation by genome scanning. Science (Washington, D.C.), **252**: 566–569.

Brutlag, D.L. 1980. Molecular arrangement and evolution of heterochromatic DNA. Annu. Rev. Genet. **4**: 121–144.

Chittenden, L.M., Schertz, K.F., Lin, Y.-R., Wing, R.A., and Paterson, A.H. 1994. A detailed RFLP map of *Sorghum bicolor* × *S. propinquum* suitable for high-density mapping suggests ancestral duplication of chromosome or chromosomal segments. Theor. Appl. Genet. **87**: 925–933.

Crowhurst, R.N., and Gardner, R.C. 1991. A genome-specific repeat sequence from kiwifruit (*Actinidia deliciosa* var. *deliciosa*). Theor. Appl. Genet. **81**: 71–78.

Deininger, P.L., Jolly, D.J., Rubin, C.M., Friedmann, T., and Schmid, C.W. 1981. Base sequence studies of 300 nucleotide renatured repetitive human DNA clones. J. Mol. Biol. **151**: 17–33.

Dvořák, J., and Zhang, H.-B. 1990. Variation in repetitive nucleotide sequences sheds light on the phylogeny of the wheat B and G genomes. Proc. Natl. Acad. Sci. U.S.A. **87**: 9640–9644.

Endrizzi, J.E., Turcotte, R.J., and Kohel, R.J. 1984. Qualitative genetics, cytology, and cytogenetics. *In* Cotton. *Edited by* R.J. Kohel and C.F. Lewis. American Society of Agronomy, Crop Science Society of America, Soil Science Society of America Publishers, Madison, WI. pp. 81–129.

Endrizzi, J.E., Turcotte, R.J., and Kohel, R.J. 1985. Genetics, cytology, and evolution of *Gossypium*. Adv. Genet. **23**: 271–375.

Ergle, D.R., and Katterman, F.R.H. 1961. DNA of cotton. Plant Physiol. **36**: 811–815.

Erlich, H.A., Gelfand, D., and Sninsky, J.J. 1991. Recent advances in the polymerase chain reaction. Science (Washington, D.C.), **252**: 1643–1651.

Evans, I.J., James, A.M., and Barnes, S.R. 1983. Organization and evolution of repetitive DNA sequences in closely related plant genomes. J. Mol. Biol. **170**: 803–826.

Fabijanski, S., Fedak, G., Armstrong, K., and Altosaar, I. 1990. A repetitive sequence probe for the C genome in *Avena* (Oats). Theor. Appl. Genet. **79**: 1–7.

Feinberg, A.P., and Vogelstein, B. 1983. A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. Anal. Biochem. **132**: 6–13.

Flavell, R.B. 1980. The molecular characterization and organization of plant chromosomal sequences. Annu. Rev. Plant Physiol. **31**: 569–596.

Flavell, R.B. 1982*a*. Repetitive sequences and genome architecture. *In* Structure and function of plant genomes. *Edited by* O. Ciferri and L. Dure. Plenum Press, New York. pp. 1–14.

Flavell, R.B. 1982*b*. Sequence amplification, deletion, and rearrangement—major sources of variation during species divergence. *In* Genome evolution. *Edited by* G.A. Dover and R.B. Flavell. Academic Press, London. pp. 301–323.

Flavell, R.B. 1986. Repetitive DNA and chromosome in plants. Philos. Trans. R. Soc. Lond. Series B, **312**: 227–242.

Flavell, R.B., Bennett, M.D., Smith, J.B., and Smith, D.B. 1974. Genome size and the proportion of repetitive nucleotide sequence DNA in plants. Biochem. Genet. **12**: 257–269.

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

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

Ganal, M.W., Lapitan, N.L.V., and Tanksley, S.D. 1988. A molecular and cytogenetic survey of major repetitive DNA sequences in tomato (*Lycopersicon esculentum*). Mol. Gen. Genet. **213**: 262–268.

Geever, R.F., Katterman, F., and Endrizzi, J.E. 1989. DNA hybridization analyses of *Gossypium* allotetraploid and two closely related diploid species. Theor. Appl. Genet. **77**: 553–559.

Gerlach, W.L., and Peacock, W.J. 1980. Chromosomal locations of highly repetitive DNA sequences in wheat. Heredity, **44**: 269–276.

Grellet, F., Delcasso, D., Panabieres, F., and Delseny, M. 1986. Organization and evolution of a higher plant alphoid-like satellite DNA sequence. J. Mol. Biol. **187**: 495–507.

Hastie, N. 1985. Middle repetitive DNA: amplification, homogeneity and mobility. Trends Genet. **1**: 37.

Hueros, G., Monte, J.V., and Ferrer, E. 1990. *Hordeum chilense* repetitive sequences: genome characterization using biotinylated probes. Theor. Appl. Genet. **80**: 24–32.

Jeffreys, A.J., Wilson, V., and Thein, S.L. 1985. Individual-specific 'fingerprints' of human DNA. Nature (London), **316**: 76–79.

Kimber, G. 1961. Basis of the diploid-like meiotic behavior of polyploid cotton. Nature (London), **191**: 98–99.

Lapitan, N.L.V. 1992. Organization and evolution of higher plant nuclear genomes. Genome, **35**: 171–181.

Ledbetter, S.A., Nelson, D.L., Warren, S.T., and Ledbetter, D.H. 1990*a*. Rapid isolation of DNA probes within specific chromosome regions by interspersed repetitive sequence polymerase chain reaction. Genomics, **8**: 475–481.

Ledbetter, S.A., Garcia-Heras, J., and Ledbetter, D.H. 1990*b*. "PCR-Karyotype" of human chromosomes in somatic cell hybrids. Genomics, **8**: 614–622.

Levy, M., Romao, J., Marchetti, M.A., and Hamer, J.E. 1991. DNA fingerprinting with a dispersed repetitive sequence resolves pathotype diversity in the rice blast fungus. Plant Cell, **3**: 95–102.

Lewin, R. 1982. Repetitive DNA still in search of a function. Science (Washington, D.C.), **217**: 621–623.

Martinez-Zapater, J.M., Estelle, M.A., and Somerville, C.R. 1986. A highly repetitive DNA sequence in *Arabidopsis thaliana*. Mol. Gen. Genet. **204**: 417–423.

Miller, D.A., Sharma, V., and Mitchell, A.R. 1988. A human-derived probe, p82H, hybridizes to the centromeres of gorilla, chimpanzee, and orangutan. Chromosoma, **96**: 270–274.

Murray, M.G., Cuellar, R.E., and Thompson, W.F. 1978. DNA sequence organization in the pea genome. Biochemistry, **17**: 5781–5790.

Murray, M.G., Palmer, J.D., Cuellar, R.E., and Thompson, W.F. 1979. Deoxyribonucleic acid sequence organization in the mung bean genome. Biochemistry, **18**: 5259–5266.

Nakamura, Y., Leppert, M., O'Connell, P., Wolff, R., Holm, T., Culver, M., Martin, C., Fujimoto, E., Hoff, M., Kumlin, E., and White, R. 1987. Variable number of tandem repeat (VNTR) markers for human gene mapping. Science (Washington, D.C.), **235**: 1616–1622.

Nelson, D.L., Ledbetter, S., Corbo, L., Victoria, D.H., and Caskey, C.T. 1989. *Alu* polymerase chain reaction: a method for rapid isolation of human-specific sequences from complex DNA sources. Proc. Natl. Acad. Sci. U.S.A. **86**: 6686–6690.

Paterson, A.H., Brubaker, C., and Wendel, J.F. 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.

Rayburn, A.L., and Gill, B.S. 1986. Isolation of a D-genome specific repetitive DNA sequence from *Aegilops squarrosa*. Plant Mol. Biol. Rep. **4**: 102–109.

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

Rivin, C.J., Cullis, C.A., and Walbot, V. 1986. Evaluating quantitative variation in the genome of *Zea mays*. Genetics, **113**: 1009–1019.

Schweizer, G., Ganal, M., Ninnemann, H., and Hemleben, V. 1988. Species-specific DNA sequences for identification of somatic hybrids between *Lycopersicon esculentum* and *Solanum acaule*. Theor. Appl. Genet. **75**: 679–684.

Singer, M.F. 1982. Highly repetitive sequences in mammalian genomes. Int. Rev. Cytol. **76**: 67–112.

Siracusa, L.D., Jenkins, N.A., and Copeland, N.G. 1991. Identification and applications of repetitive probes for gene mapping in the mouse. Genetics, **127**: 169–179.

Stallings, R.L., Torney, D.C., Hildebrand, C.E., Longmire, J.L., Deaven, L.L., Jeff, J.H., Dogget, N.A., and Moyzis, R.K. 1990. Physical mapping of human chromosomes by repetitive sequence fingerprinting. Proc. Natl. Acad. Sci. U.S.A. **87**: 6219–6222.

Thompson, W.F., and Murray, M.G. 1981. The nuclear genome: structure and function. *In* The biochemistry of plants. Vol. 6. *Edited by* A. Marcus. Academic Press, New York. pp. 1–81.

Vanderwiel, P.S., Voytas, D.F., and Wendel, J.F. 1993. Copia-like retrotransposable element evolution in diploid and polyploid cotton (*Gossypium* L.). J. Mol. Evol. **36**: 429–447.

Walbot, V., and Dure, L.S. 1976. Developmental biochemistry of cotton seed embryogenesis and germination. VII. Characterization of the cotton genome. J. Mol. Biol. **101**: 503–536.

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

Wilson, J.T., Katterman, F.R.H., and Endrizzi, J.E. 1976. Analysis of repetitive DNA in three species of *Gossypium*. Biochem. Genet. **14**: 1071–1075.

Zamir, D., and Tanksley, S.D. 1988. Tomato genome is comprised largely of fast evolving low copy number sequences. Mol. Gen. Genet. **129**: 254–261.

Zhao, X., and Kochert, G. 1992. Characterization and genetic mapping of a short, highly repetitive, interspersed DNA sequence from rice (*Oryza sativa* L.). Mol. Gen. Genet. **231**: 353–359.

Zhao, X., and Kochert, G. 1993*a*. Phylogenetic distribution and genetic mapping of $(GGC)_n$ microsatellite from rice (*Oryza sativa* L.). Plant Mol. Biol. **21**: 607–614.

Zhao, X., and Kochert, G. 1993*b*. Clusters of interspersed repetitive DNA sequences in the rice (*Oryza*). Genome, **36**: 944–953.

Zhao, X., Wu, T., Xie, Y., and Wu, R. 1989. Genome-specific repetitive sequences in the genus *Oryza*. Theor. Appl. Genet. **78**: 201–209.