Plant Molecular Biology **38:** 1031–1042, 1998.

Macromolecular organization and genetic mapping of a rapidly evolving chromosome-specific tandem repeat family (B77) in cotton (*Gossypium*)

Xinping Zhao[1], Yuanfu Ji[1], Xiaoling Ding[1], David M. Stelly and Andrew H. Paterson*
Department of Soil and Crop Sciences, Texas A& M University, College Station, TX 77843-2474, USA (author for correspondence); University of Michigan Medical Center, 5110 CCGC, 1500 E. Medical Center Dr., Ann Arbor, MI 48109-0940, USA*

Received 30 October 1997; accepted in revised form 4 June 1998

Key words: evolution, fluorescence *in situ* hybridization (FISH), genome, repetitive DNA

Abstract

Isolation and characterization of the most prominent repetitive element families in the genome of tetraploid cotton (*Gossypium barbadense* L; [39]) revealed a small subset of families that showed very different properties in tetraploids than in their diploid progenitors, separated by 1-2 million years. One element, B77, was characterized in detail, and compared to the well-conserved 5S and 45S rRNA genes. The 572 bp B77 repeat was found to be concentrated in several discontinuous tandem arrays confined to a single 550 kb *Sal*I fragment in tetraploid cotton. Genetic mapping based on the absence of the pentameric 'rung' in the *G. barbadense* 'ladder' showed that B77 maps to a D-subgenome chromosome. *In situ* hybridization supports the contention that the array is confined largely to a single chromosomal site in the D-subgenome. The B77 repeat has undergone a substantial increase in copy number since formation of tetraploid cotton from its diploid relatives. RFLPs observed among tetraploid cotton species suggest that amplification and/or rearrangement of the repeat may have continued after divergence of the five tetraploid cotton species. B77 contains many short direct repeats and shares significant DNA sequence homology with a *Nicotiana alata* retrotransposon Tna1-2 integrase motif. The recent amplification of B77 on linkage group D04 suggests that the D-subgenome of tetraploid cotton may be subject to different evolutionary constraints than the D-genome diploid chromosomes, which exhibit few genome-specific elements. Further, the abundance of B77 in *G. gossypioides* supports independent evidence that it may be the closest extant relative of the D-genome ancestor of cotton.

Introduction

The genus *Gossypium* L. has long been a focus of genetic, systematic, and breeding research. *Gossypium* is comprised of about 50 diploid and tetraploid species indigenous to Africa, Central and South America, Asia, Australia, the Galapagos, and Hawaii [13, 14]. Diploid species are all $n = 13$, and fall into 7 different 'genome types', designated A through G based on chromosome pairing relationships [3, 10, 11]. A total of 5 tetraploid ($n = 2x = 26$) species are recognized. All tetraploid cottons exhibit disomic chromosome pairing [22]. 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 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). The two tetraploid species dominate world cotton production, and a large number of improved varieties have been developed.

As part of a long-term study of the organization and evolution of the cotton genome, we have

The nucleotide sequence data reported will appear in the GenBank/EMBL Nucleotide Sequence Databases under the accession number AF025353.

developed a detailed molecular map of the cotton chromosomes [29], and have cloned and characterized the majority of repetitive elements from tetraploid cottons [39]. Both the repertoire and chromosome organization of low-copy DNA is largely conserved between A and D subgenomes, however repetitive DNA comprises a large and rapidly-evolving portion of the cotton genome [42]. Considerable variation is found among the set of repetitive DNA element families present in different diploid cottons. However, all families are found in all tetraploid cottons [42], consistent with a recent origin from a single polyploidization event.

We report here the isolation and characterization of a rapidly-evolving tandem repeat family (B77) from tetraploid cotton. There appear to have been 3 waves of amplification of the B77 family since the divergence of *Gossypium* and *Thespesia* from a common ancestor. The third wave has produced a concentration of elements that occur in clusters of discontinuous tandem arrays on a D-subgenome chromosome of tetraploid cotton. The chromosomal location of this repeat is confirmed by both genetic mapping and *in situ* hybridization. The D-subgenome of tetraploid cotton may be subject to different evolutionary constraints than the D-genome diploid chromosomes, which exhibit few genome-specific elements.

Materials and methods

Plants and DNA isolation

Cotton genotypes used in genomic Southern blot analysis, and in slot blot hybridization analysis of copy number and genome specificity, are listed in Table 1. Genomic DNA was isolated according to Paterson *et al.* [28]. High-molecular-weight (HMW) DNA was isolated according to Zhao *et al.* [41].

Library construction and repetitive DNA sequence selection

Plasmid library construction

Genomic DNA from *G. barbadense* 'Pima S6' was digested with *Bam*HI and DNA fragments were separated by electrophoresis in a 1% agarose gel. DNA fragments of 0.5-2 kb were isolated from the gel, ligated into the *Bam* HI site of pGEM11Zf (t) (Promega), and transformed into *Escherichia coli* strain DH5a. Recombinants were selected on LB plates containing ampicillin, X-gal and IPTG. Inserts of recombinants were PCR-amplified with M13 universal primers as described elsewhere [5], then electrophoresed (1% agarose) and blotted onto nylon membranes (Hybond N^+). The membrane-bound DNA was hybridized to ^{32}P-dCTP-labeled Pima S6 genomic DNA and tobacco chloroplast DNA. The probes were prepared using random primer oligolabeling [12]. Putative nuclear repeated DNA sequences were identified by strong hybridization to Pima S6 genomic DNA, but not to tobacco chloroplast DNA.

18S rDNA (pXP108) was isolated from a genomic library described elsewhere [39]. It contains a 265 bp DNA fragment showing > 95% homology to the *G. hirsutum* 18S rRNA gene (positions 339 to 604) [24] (data not shown). 5S rDNA (R60) was initially isolated from a *G. raimondii* genomic library made in the same way as described [39]. Additional clones were isolated from a genomic library of *G. hirsutum* 'TX9' *Bam*HI fragments in pBluescript KS (+). One clone, R60-11, contains the full-length 5S rRNA gene and was used in this study. The DNA sequence of R60-11 is identical to that of 'hirsutum 9' sequenced previously [7].

Lambda library construction and screening

HMW DNA from *G. barbadense* 'K101' was isolated as described [41], partially digested with *Sau*3AI, and fractionated by electrophoresis on a 1% low-melting agarose gel. Fragments of 10–23 kb were ligated into the *Bam*HI site of lambda dash II (Stratagene), and packaged with Gigapack II packaging extract (Stratagene) following the manufacturer's instructions. Recombinants were selected by propagation in *E. coli* strain p2392. The library of 200 000 primary recombinants, with an average size of 15 kb, was amplified once according to the manufacturer's suggestions.

To isolate more B77-related sequences, we screened the amplified library using purified insert of the B77 plasmid as a hybridization probe. DNA was isolated from positive clones, digested with a series of restriction endonucleases, and Southern blotted. The membranes were then hybridized to the ^{32}P-dCTP-labeled B77 insert. Three individual clones were selected that contain different insert sizes and were used to study the organization of B77 in the cotton genome.

Fluorescence in situ *hybridization (FISH)*

Mitotic and meiotic chromosome preparations followed the procedures of Jewell and Islam-Faridi [21] and Crane *et al.* [6], respectively. FISH with biotiny-

Table 1. Cotton cultivars and species.

Number	Cultivars	Species	Genome group
Cultivars for Southern blot analysis (for Figure 5)			
1.	Pee Dee 2164	*G. hirsutum*	AD
2.	Delta Pine 50	*G. hirsutum*	AD
3.	Stoneville 213	*G. hirsutum*	AD
4.	Tamcot	*G. hirsutum*	AD
5.	Westburn-M	*G. hirsutum*	AD
6.	CA 3029	*G. hirsutum*	AD
7.	Paymaster HS 26	*G. hirsutum*	AD
8.	Acala 1517-75	*G. hirsutum*	AD
9.	Acala	*G. hirsutum*	AD
10.	Acala Maxxa	*G. hirsutum*	AD
11.		*G. lanceolatum* race *palmeri*	AD
12.	K101	*G. barbadense*	AD
13.	AZK263	*G. barbadense*	AD
Species for Southern blot analysis			
1.	TM1	*G. hirsutum*	AD
2.		*G. lanceolatum race palmeri*	AD
3.	K101	*G. barbadense*	AD
4.	PW 44	*G. darwinii*	AD
5.	GH 3547	*G. tomentosum*	AD
6.	AD 4-7	*G. mustelinum*	AD
7.		*G. herbaceum*	A
8.		*G. arboreum*	A
9.		*G. raimondii*	D
10.		*G. trilobium*	D
11.		*G. sturtianum*	C
12.		*G. longicalyx*	F
13.		*Hibiscus calyphyllus*	Outgroup
14.		*Thespesia lampas*	Outgroup

lated probe B77 followed the procedures described previously [18]. After initial FISH with B77, the same preparations were washed and reprobed with biotinylated A_2 genome DNA as described [19] with some modifications. Briefly, the slides were washed twice in detection buffer at room temperature for 30 min each, and then washed twice in 2 × SSC at 37 °C for 5 min each to remove the anti-fade solution. The slides were then reprobed with biotinylated A_2 genome DNA as described [20].

Genetic mapping

Genetic mapping used a well-characterized population derived from a cross between *G. barbadense* L. and *G. hirsutum* L., and analytical techniques previously described [29].

DNA sequence analysis

Cloned plasmid DNA was used for sequencing by the dideoxynucleotide termination method [32]. Both strands were sequenced using a Sequenase Version II sequencing kit (USB) with ^{35}S-dATP and both the M13 universal and reverse primers.

Southern blot hybridization analysis

Five micrograms of 'Pima S6' genomic DNA were digested with 18 restriction enzymes (*Alu*I, *Bam*HI, *Bgl*I, *Bst*NI, *Cfo*I, *Dra*I, *Eco*RI, *Eco*RV, *Hae*III, *Hha*I, *Hind*III, *Hpa*II, *Pst*I, *Pvu*II, *Sau*3AI, *Sfi*I, *Taq*I, and *Xba*I) for investigating the organization of repeated DNA sequences in the cotton genome. Genomic DNA samples from different genotypes were digested with 8 restriction enzymes for investigating variation across

Table 1. Continued. Cotton cultivars and species

Number	Cultivars	Species	Genome group
Species for slot blot analysis (for Figure 4)			
1.	#123	*G. aridum*	D
2.	D9-3	*G. laxum*	D
3.	D10-9	*G. thurberi*	D
4.	D8-4	*G. trilobum*	D
5.	D3k55	*G. klotzschianum*	D
6.	#32A	*G. davidsonii*	D
7.	D5-37	*G. raimondii*	D
8.	D6-2	*G. gossypioides*	D
9.	A2-47	*G. arboreum*	A2
10.	JMS	*G. herbaceum*	A1
11.	TX9	*G. hirsutum*	AD
12.	K101	*G. barbadense*	AD
13.		*G. anomalum*	B
14.		*G. triphyllum*	B
15.		*G. robinsonii*	C
16.	AZ40	*G. sturtianun* (var. *nandewarense*)	C
17.	E2	*G. somalense*	E2
18.		*G. longicalyx*	F
19.	G1-4	*G. bickii*	G
20.		*Thespesia lampas*	outgroup

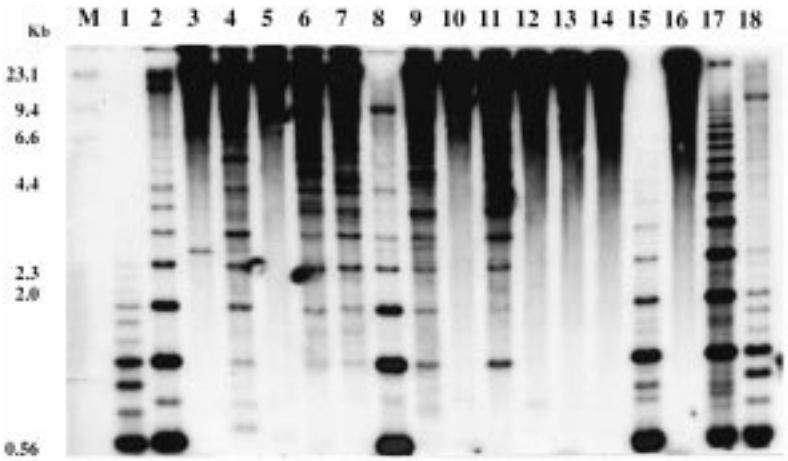

Figure 1. Southern blot analysis showing the tandem organization (A, left) and methylation (B, right) of the B77 sequence in the tetraploid cotton genome. A. Lanes 1–18: 'Pima S6' genomic DNA digested with *Alu*I, *Bam*HI, *Bgl*I, *Bst*NI, *Cfo*I, *Dra*I, *Eco*RI, *Eco*RV, *Hae*III, *Hha*I, *Hind*III, *Hpa*II, *Pst*I, *Pvu*II, *Sau*3AI, *Sfi*I, *Taq*I, and *Xba*I, respectively. M stands for lambda molecular marker digested with *Hind*III. B. 'Pima S6' genomic DNA digested with a pair of isoschizomeric restriction endonucleases differring in methylation sensitivity (*Sau*3AI and *Mbo*I). Molecular sizes are indicated to the very left.

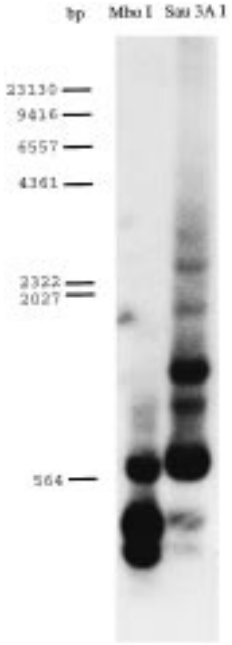

Figure 1. B.

different species and for evaluating their potential use as DNA fingerprinting probes for identification of cultivated cotton varieties. Southern blotting and hybridization were carried out as described elsewhere [5].

Slot blot hybridization analysis

Genomic DNA from representatives of the seven genome groups of *Gossypium*, two tetraploid species, and *Thespesia lampas* (Table 1) was used in slot blot hybridization analysis according to Zhao *et al.* [39]. One microgram of DNA from two tetraploid cottons and 0.5 mg of DNA for the other species were used. Three sets of slot blots were prepared, and hybridized to the B77 sequence, 5S rDNA, and 18S rDNA respectively as described [40]. After being washed in a solution of 0.5 × SSC/0.1% SDS at 65 °C for 30 min and exposed to X-ray films, the blots were further washed in 0.1 × SSC/0.1% SDS at 70 °C for 30 min and re-exposed. Quantitative analyses of slot blot hybridization were carried out as described earlier [40]. The relative signal intensities (Figure 4) were adjusted for the genome size using multipliers of 0.35 for D-genome, and 0.65 for A, E, F, and B genomes, based on direct measurement (H. J. Price and S. Johnston, pers. comm.), and 0.8 for C genome and 0.7 for the G genome based on ratios of DNA content in the respective diploids to that of the tetraploid [9].

CHEF analysis

HMW DNA from *G. bardadense* 'K101' was digested with various restriction enzymes and electrophoresed on a 1% agarose gel in 0.5× TBE as described [41]. The DNA in the gel was nicked with 60 mJ of UV light (254 nm) using the GS Gene Linker (BioRad), transferred onto a Hybond N+ membrane (Amersham, USA) using a blotting solution (0.4 M NaOH, 1.5 m NaCl) for 40 h, then hybridized to the B77 and 5S rDNA sequences as described [5].

Results

Organization of B77, 5S rDNA and 45S rDNA families in diploid and tetraploid cottons

Complete digests of genomic DNA of *G. barbadense* 'Pima S6' displayed the 'ladder' typical of tandemly repeated sequences in *Bam*HI, *Eco*RV, *Sau*3AI, and *Taq*I digests (Figure 1A, lanes 2, 8, 15, and 17). The ladder suggested a length of about 570 bp for the B77 monomer. In all cases where ladders are evident, there were some weakly-hybridizing bands that are not multimers of the basic repeat unit, suggesting that the B77 family either is not completely homogeneous or has some homology with other DNA sequences in the genome. Similarly, the tandem organization for both 5S and 45S rDNA was also observed by probing the same blot with R60-11 and pXP108 respectively (data not shown).

Methylation was investigated by comparing hybridizations on digests with isoschizomeric restriction enzymes differing in methylation sensitivity (*Sau*3AI and *Mbo*I) as described previously [39]. *Sau*3AI and *Mbo*I were then used because there are two recognition sites for the enzyme pair within the B77 repeat (see Figure 8). *Mbo*I (methylation-insensitive) showed much more complete digestion that *Sau*3AI (methylation-sensitive), suggest that the B77 repeat is methylated (Figure 1B). 5S and 45S rDNA are moderately methylated (data not shown).

To investigate the long-range organization of the B77 sequence in tetraploid cotton, (HMW) DNA was isolated, restriction-digested, separated by CHEF, and Southern-blotted. B77 detected a single *Sal*I DNA fragment of about 550 kb (Figure 2. lane 4). A DNA fragment of this size could contain a maximum of 900 copies of the B77 array, very close to the copy number we estimated by quantitative slot blot hybridization as described elsewhere [39]. Multiple hybridization

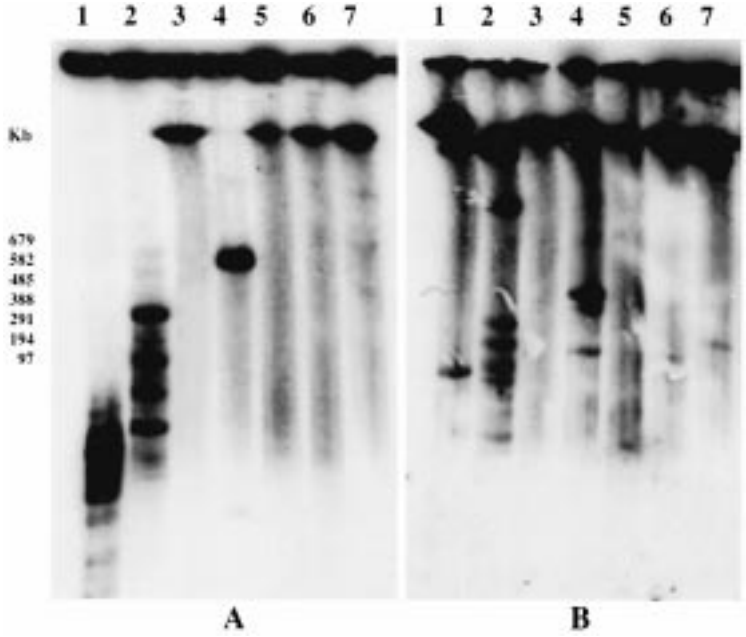

Figure 2. Southern blot hybridization of the CHEF gel to the B77 sequence (A) and 5S rDNA (B) showing long range organizations of the repeats in the tetraploid cotton genome. *G. barbadense* 'K101' HMW DNA were digested with 40 units of different restriction enzymes (lanes 1–7: *Nco*I, *Nhe*I, *Not*I, *Sal*I, *Sfi*I, *Sma*I, and *Pvu*II, respectively). Lambda ladders (FMC Scientific) were used as molecular weight markers. Digested HMW DNA and lambda markers were loaded in a 1% agarose and fractionated by pulsed field gel electrophoresis (PFGE) in 0.5× TBE (1× TBE is 89 mM Tris-borate, 2 mM EDTA) buffer using the HEXCHEF 6000 (CBS-Scientific) set with the MJ Research Programmable Power Inverter: A = 100 s; B = −0.05 s; C = 100 s; D = −0.05 s; E = 175 timed; F = −0.01 sec; and G = −0.01 s. The CHEF gel was run for 60= h with 150 V at 11 °C, Southern-blotted, and hybridized to the B77 sequence. The same blot was stripped and rehybridized to 5S rDNA.

bands on *Nco*I and *Nhe*I (Figure 2, lane 1 and 2) digests suggested that the 550 kb DNA fragment is not comprised exclusively of the B77 sequence, since these enzymes do not have restriction sites in the B77 monomer (see Figure 8). Alternatively, members of B77 family may not be completely uniform in sequence and a few repeat units may contain recognition sites for these enzymes. For other rare-cutting enzymes (*Not*I, *Sfi*I, *Sma*I, and *Pvu*II; Figure 2, lanes 3, 5, 6, and 7, respectively), B77 hybridized to the unresolved region of the CHEF gel.

While genomic digests with most rare-cutting restriction enzymes showed only a single restriction fragment when hybridized with B77, the same blot probed with a full-length 5S repeat (R60-11; Figure 2B) revealed two bands, consistent with the results of *in situ* hybridization showing that 5S arrays occupy a single chromosomal location in each of two subgenomes in AD genome allopolyploids [6].

To test the degree of continuity of the B77 array, a lambda genomic library was screened with the B77 insert, and three clones were selected. Restriction mapping indicated that these three lambda clones do not consist of the B77 sequence exclusively (Figure 3), suggesting that the B77 array may be interrupted in the genome.

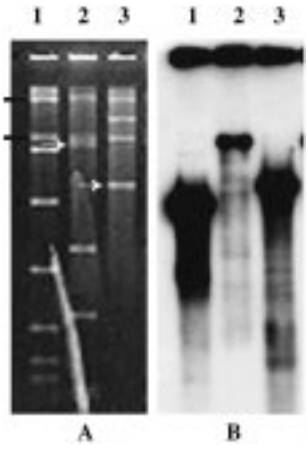

Figure 3. Restriction enzyme digestion and Southern hybridization patterns of B77-containing lambda phage clones. DNA from three lambda clones were digested with *Eco*RI (A), blotted, and hybridized to B77 sequence (B). Vector arms are labeled with a line and DNA fragments hybridized to B77 are indicated with arrows in A.

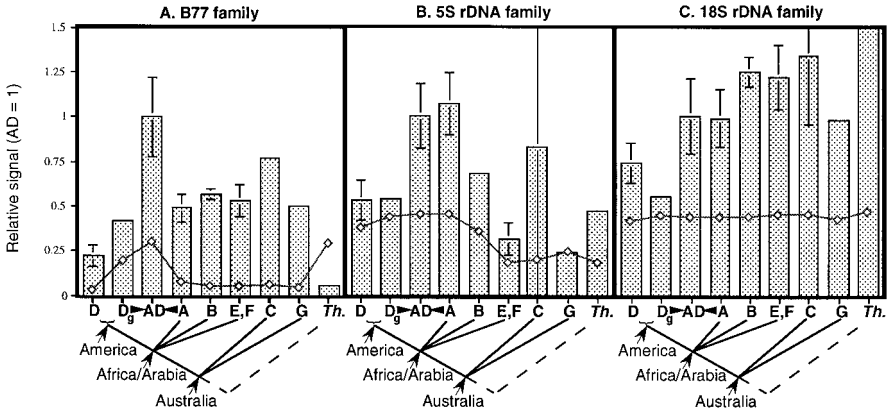

Figure 4. Quantitative slot blot analysis showing the distribution of the B77 sequence (A), 5S rDNA (B), and 18S rDNA (C) in species of *Gossypium* and the outgroup *Thespesia lampas*. The slot blots were washed at 0.5× SSC/0.1% SDS at 65 °C (bars), and further washed at 0.1× SSC/0.1% SDS at 70 °C (curves) before autoradiographed. DNA samples used are listed in Table 1.

The hybridization patterns of B77 in diploid cottons are quite different from those of tetraploids. Instead of ladders of hybridization bands on Bam HI digests as seen for tetraploid cottons, diploids produced hybridization signals corresponding to much higher molecular weight regions (not shown), suggesting different organization or methylation of the repeat family between tetraploid and diploid cottons. Different patterns of hybridization were observed for 7 additional restriction enzyme digests between tetraploid and diploid cottons (data not shown).

Phylogenetic distribution of B77

Genomic DNA from representatives of each of the seven diploid *Gossypium* genomes (A to G) as well as *Thespesia lampas* (see Table 1) were hybridized to the B77 sequence, 5S rDNA (R60-11), and 18S rDNA (pXP108), separately, in slot blot analysis.

All *Gossypium* species contain the B77 sequence, but vary considerably in copy number. Quantitative analysis of the slot hybridization signals shows that signal per genome equivalent varied by 17-fold for the B77 sequence, 7-fold for 5S rDNA, and 3-fold for 45S rDNA (Figure 4). B77 hybridization signals in the tetraploid cottons were twice that of either the A or D-genome progenitor diploid genomes (per genome equivalent), as well as all other *Gossypium* genomes. Among diploid *Gossypium* genomes, there is 2- to 3-fold difference in hybridization signal (per genome equivalent) from the B77 sequence. In contrast, 45S (18S) rDNA hybridized to all species with similar intensity (Figure 4C). 5S rDNA showed more variation in copy number than 18S rDNA, but less variation than B77 (Figure 4B).

An outgroup, *T. lampas* showed discernible, but weak, hybridization to B77. For 5S and 18S rDNA, *T. lampas* produced hybridization signals similar to the *Gossypium* species. The *c*-value of *T. lampas* is unknown; if it were conservatively estimated to be as small as that of the D genome of *Gossypium*, then *Thespesia* would have 10- to 31-fold fewer copies of B77 than *Gossypium* genomes. Even if it were estimated to be the average genome size of *Gossypium*, *Thespesia* would still have 6 to 19-fold fewer copies of B77 than *Gossypium* genomes.

To further characterize DNA sequence variation of the B77 family within individual taxa, we compared signal after washes at two different stringencies (0.5 × SSC/0.1% SDS at 65 °C for 30 min, versus 0.1 × SSC/0.1% SDS at 70 °C for 30 min (Figure 4). The 18S rDNA family provides a standard by which to interpret our results, retaining 39.4% to 45.6% of hybridization signals after the more stringent wash across all taxa. The 5S rDNA was very similar to the 18S rDNA, retaining more than 33% of hybridization signals in tetraploid cotton and its immediate progenitors. However in the more distantly related E and F genomes, as well as Australian C and G genomes, and the outgroup (*Thespesia*), 5S rDNA signal showed a marked drop after high-stringency wash.

The B77 family is rapidly evolving in tetraploid cottons

A set of cotton cultivars representing germplasm cultivated in the USA were analyzed using the B77

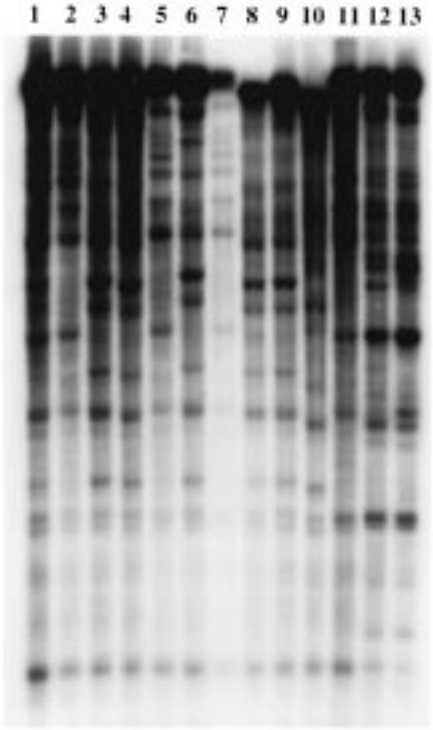

Figure 5. Variety-specific RFLPs generated by probing with the B77 sequence on *Dra*I digests of DNA samples (see Table 1 for detail).

sequence as a probe in Southern blots of *Dra*I digests. Most cultivars show a unique RFLP profile (Figure 5). Other restriction enzyme digests (data not shown) show a similarly high level of variation. These results suggest that B77 continues to be rapidly evolving, and indicate that the B77 repeat might be used as a DNA fingerprinting probe for cotton cultivar identification.

Chromosomal location of the B77 family

The observation that the B77 repeat is largely confined to a single 550 kb *Sal*I DNA fragment suggests that it has undergone differential amplification in the two subgenomes of tetraploid cotton. *In situ* hybridization experiments further support this hypothesis. FISH of B77 yields signal on just one chromosome pair. Chromosome painting indicates that the locus is on a D-subgenome chromosome (Figure 6).

Genetic mapping reveals the locus of the B77 array. The pentameric rung in the B77 ladder was absent from *G. barbadense* accession 'K101', used in the primary mapping population of cotton. Segregation for presence or absence of this rung permitted RFLP mapping of B77 to the present terminus of a linkage group designated 'D04' [29]. Prior mapping of duplicated loci detected by individual low-copy DNA probes shows that linkage group D04 is homoeologous to cotton chromosome 10 of the A-subgenome (Figure 7).

DNA sequence analysis of the B77 family

The B77 plasmid contains a 572 bp insert, with a GC content of 40% (Figure 8). The sequence contains many internal direct repeats of 8 to 12 bp in length (only direct repeats of 12 bp or more are shown in Figure 8). A short open reading frame is found in the sequence (1–138). GenBank search shows that a region of bases 40–264 in the B77 repeat shows 60% homology to *Nicotiana alata* retrotransposon Tna1-2 integrase motif (a region from 911-1135) [31]. However, the Tna1-2 element contains no open reading frame and is unlikely to be functional in transposition. A pair of direct repeats (12 bp) are located about 10 bp away from the boundaries of the homologous region (49-60 and 243–254) (Figure 8).

Discussion

Our interest in B77 was stimulated by its increase in copy number in the 1–2 million years since the formation of polyploid *Gossypium* from diploid ancestors [34]. B77 is exceptional in this regard, as most other repetitive DNA families in polyploid cotton are present in equal or fewer total copies than the sum of the copy numbers in the diploid progenitors (Zhao *et al.*, submitted). For example, the signal for 5S and 18S rDNAs (Figure 4) in the tetraploids is similar to that of the A-genome diploid alone.

Both B77 and other repetitive DNAs in cotton underline the dynamic nature of repetitive sequence evolution. For example, despite minimal net change in copy number, we have shown that other cotton repetitive element families have moved to new locations since polyploid formation (Zhao *et al.*, submitted). By contrast, however, B77 appears to have undergone amplification primarily in a genomic region small enough to behave essentially as a single genetic locus.

A possible model for B77 evolution

By considering jointly the hybridization intensity of moderate stringency slot blots (reflecting copy number), and the signal retained after stringent washes (reflecting family heterogeneity), the taxa studied fall

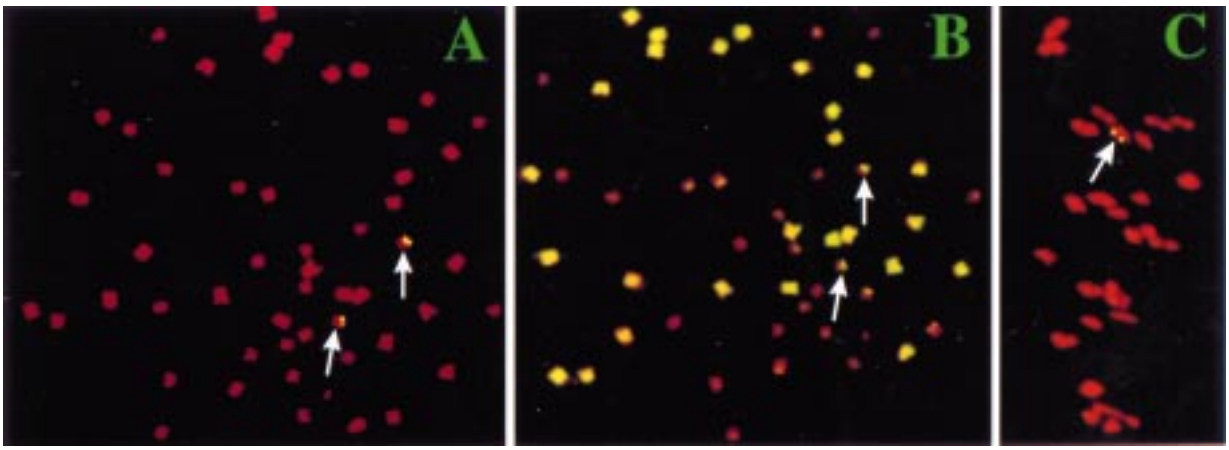

Figure 6. Fluorescence *in situ* hybridization (FISH) of the tetraploid cotton *G. hirsutum* to the B77 sequence showing that the repeat is located in one pair of the D chromosomes. (A) Root-tip metaphase cell spread after FISH showing that a single pair of chromosomes bear hybridization sites of B77 (arrows), detected by yellow fluorescein isothiocyanate (FITC). (B) Genomic*in situ* hybridization (GISH): A-subgenome chromosomes were 'painted' as yellow after reprobing the same spread from (A) with biotinylated A2-genome DNA probe (blocked with D5 genomic DNA), and the pair of B77 hybridization sites (arrows) were shown to be on one pair of D-subgenome chromosomes. (C) Meiotic metaphase I cell spread of a *G. hirsutum* translocation heterozygote (NT9L-17R), showing that just one bivalent bears the B77 FISH signals (arrow).

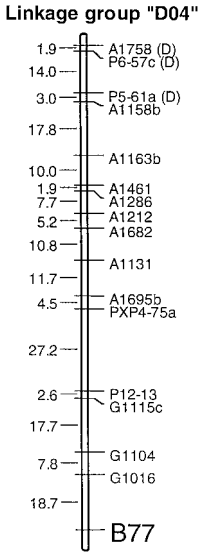

Figure 7. Chromosomal locations of the B77 repeat in the cotton RFLP map using a well-characterized cotton mapping population [29].

into three distinct classes that suggest a possible model for the evolution of the B77 element.

1. Tetraploid cottons and *G. gossypioides* (their possible progenitor [36]; Zhao *et al.*, submitted) showed relatively less sequence divergence, as reflected by retention of 28% and 19% (respectively) of B77 signal at high stringency. Tetraploids show the highest copy number (signal intensity) of B77, and *G. gossypioides* has about twice the copy number of the other D-genome taxa. This class appears to have relatively high abundance of B77, and relatively low heterogeneity, suggesting recent amplification.

2. All diploid taxa except *G. gossypioides* showed a relatively high degree of sequence divergence among their B77 families, as reflected by retention of only 2.6–6.7% of hybridization signals at high stringency. The high heterogeneity of B77 family members in this class is accompanied by low abundance, suggesting ancient amplification followed by divergence.

3. *Thespesia lampas*, although containing only about 10% of the copy number of the tetraploids for B77, showed uniformity of B77 sequences at a level matched only by the tetraploids, with about 28% of signal retained after the high-stringency wash. *T. lampas* appears to contain only a small number of B77 copies that are relatively homogeneous.

The occurrence of these three classes suggests that B77 has undergone three waves of amplification, in the course of *Gossypium* evolution. The first wave of amplification in the *Gossypium* taxon may have preceded radiation of most diploid *Gossypium* genomes from a common ancestor, causing higher levels of B77 in the A, C, D (except for *G. gossypioides*, see below), E, F, and G genomes than *Thespesia lampas*. A second wave appears to have occurred specifically in the Australian C genome after it diverged from other diploid taxa. While we have studied only two C-genome taxa, both exhibited similarly high B77 copy number (error bars on the graph are at ± 0.01, too small

Figure 8. DNA sequence of the B77 repeat. Short tandem repeats ($\geq$ 12 bp) are labeled as pairs underneath the nucleotide sequences.

to discern). Moreover, these two C-genome taxa exhibit marked differences from each other in the levels of many other repetitive element families (Zhao *et al.*, submitted), suggesting that the similarity in B77 levels is noteworthy.

To account for the high degree of heterogeneity among B77 families in diploid cottons, each of the first two waves of amplification must be relatively ancient, surely predating the divergence of the A and D diploid genomes estimated to have been about 10 million years ago [34, 35]. The third wave is recent, during or after formation of tetraploid *Gossypium* from A and D genome ancestors about 1–2 million years ago, accounting for both the high copy number and the lower sequence divergence among B77 family members found in tetraploid cottons. The remarkably high degree of homogeneity in *T. lampas* may suggest that its few B77 copies have been subject to constraints on their divergence, perhaps as a result of functional importance.

The finding that the B77 family of tetraploid cotton is largely confined to a single region of a D subgenome chromosome, despite the paucity of B77 elements in most D genome diploid cottons, may suggest that the D subgenome of tetraploid cotton is subject to different evolutionary constraints than the D genome of diploid cottons.

While we suggest that there have been three major waves of B77 amplification, the observation that B77 produced unique RFLP profiles on many closely related cultivated cottons shows that B77 is not static between waves but evolves continuously, albeit perhaps at variable rates.

B77 and Gossypium gossypioides

Both genetic mapping and *in situ* hybridization show that most B77 family members in tetraploid cotton occur in a single discontinuous block on a D-subgenome chromosome (linkage group D04). However, most diploid D-genome taxa contain a paucity of B77 repeats, consistent with the patterns of distribution for many other repeat families. Among the diploid cotton genomes, D is smallest, with a *c*-value estimated at 1.01 pg. Most repetitive DNA elements in the tetraploid (AD) are abundant in the ancestral A genome, but rare/absent in the ancestral D genome. In fact, 'A-genome-specific' and 'A-genome-enriched' repeats account for at least half of the difference in *c*-value between A and D genomes (Zhao *et al.*, submitted). Both regarding the 'A-genome-enriched' repeats and B77, modern D genome taxa contain even fewer copies than modern C and G genomes. Yet, formation of a distinct C/G lineage is thought to far predate divergence of the D, A, B, E, and F genomes (see Figure 4).

The evolution of B77 reinforces several other lines of evidence that associate *G. gossypioides* with formation of polyploid cotton. Both the copy number, and the relatively low heterogeneity of B77 family members in *G. gossypioides* much more-closely resembles that of the tetraploids than other D-genome diploids (Figure 4). These similarities between *G. gossypioides* and tetraploid cottons are consistent with several other lines of evidence including ribosomal DNA sequence [37], and interspersed repetitive DNAs (Zhao *et al.*, submitted) which suggest that *G. gossypioides* may be either the closest living descendant of the New

World cotton progenitor, or a relic of polyploid cotton formation.

Genetic mapping of tandemly repetitive DNAs facilitates the integration of genetic and physical maps

In plants, many tandem repeats have been localized to specific chromosomal regions such as cemtromere, telomere, or heterochromatin by *in situ* hybridization, making them excellent landmarks for studying chromosome structure, function, and evolution. However, few tandem repeats (excluding microsatellites) have been genetically mapped [15, 30, 38]. Recently, tandem repeats have been used in genetic and cytogenetic mapping in pea [16, 17]. Inclusion of repetitive DNA elements will increase the utility of the cotton genetic map [29] in both basic and applied research, facilitating the integration of genetic and physical maps.

The B77 repeat mapped to the present terminus of linkage group D04 [29], which is homoeologous to cotton chromosome 10. FISH analysis suggests that, although it is localized toward the telomeric region of the chromosome, the B77 repeat is clearly not located at the telomere (Figure 6). These results suggest that the current cotton linkage group [29] may not yet reflect the full chromosome. The B77 repeat is nearly 20 cM from the nearest DNA marker, in a region where few markers are mapped. Repetitive DNA elements may help fill gaps in existing genetic maps, in genomic regions that are under-populated with low-copy DNA sequences.

Toward assembly of a set of chromosome-specific repeats for the cotton genome

B77 (specific to LG D04), together with the 5S (specific to chromosomes 9 and 23) and 18S ribosomal DNAs (specific to chromosomes 9, 16, and 23) [6], provide the foundation for a set of chromosome-specific repeats in cotton. Chromosome-specific repeats are useful for diagnosis of deletions, translocations, and insertions, and have been isolated from various human chromosomes including the Y [4, 8, 23, 27, 33], as well as from mouse [25, 26] and *Drosophila* [2]. Chromosome-specific repeats are also useful for flow-sorting and assembly of chromosome-specific DNA libraries [1]. Together with B77, 6 of 26 cotton chromosomes can now be identified using these specific probes.

Acknowledgments

We thank Dr Vipin Rastogi for technical assistance; and Dr Shirley Coomber for providing tobacco chloroplast DNA. Aspects of this work were supported by grants from the Texas Higher Education Coordinating Board and USDA Plant Genome Research Program, to A.H.P. and D.M.S..

References

1. Arumuganathan K, Martin GB, Telenius H, Tanksley SD, Earle ED: Chromosome 2-specific DNA clones from flow-sorted chromosomes of tomato. Mol Gen Genet 242: 551–558 (1994).
2. Awgulewitsch A, Bunemann H: Isolation of Y-chromosomal repetitive DNA sequences of *Drosophila hydei* via enrichment of chromosome-specific sequences by heterogeneous hybridization between female and male DNA. J Biochem Biophys Meth 12: 37–50 (1986).
3. Beasley JO: Meiotic chromosome behavior in species hybrids, haploids, and induced polyploids of*Gossypium*. Genetics 27: 25–54 (1942).
4. Burk RD, Stamberg J, Young KE, Smith KD: Use of repetitive DNA for diagnosis of chromosomal rearrangements. Hum Genet 64: 339–342 (1983).
5. Chittenden LM, Schertz KF, Lin Y-R, Wing RA, Paterson AH: 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 (1994).
6. Crane CF, Price HJ, Stelly DM, Czeshin DG: Identification of a homoeologous chromosome pair by in situ DNA hybridization to ribosomal RNA loci in meiotic chromosomes of cotton (*Gossypium hirsutum*). Genome 35: 1015–1022 (1993).
7. Cronn EC, Zhao X, Paterson AH, Wendel JF: Polymorphism and concerted evolution in a tandemly repeated gene family: 5S ribosomal DNA in diploid and allopolyploid cottons. J Mol Evol 42: 685–705 (1996).
8. Disteche C, Luthy D, Haslam DB, Hoar D: Prenatal identification of a deleted Y chromosome by cytogenetics and a Y-specific repetitive DNA probe. Hum Genet 67: 222–224 (1984).
9. Edwards GA, Endrizzi JE: Cell size, nuclear size, and DNA content relationships in *Gossypium*. Can J Genet Cytol 17: 181–186 (1976).
10. Endrizzi JE, Turcotte RJ, Kohel RJ: Qualitative genetics, cytology, and cytogenetics. ASA/CSSA/SSSA Publishers, Madison, WI (1984).
11. Endrizzi JE, Turcotte RJ, Kohel RJ: Genetics, cytology, and evolution of *Gossypium*. Adv Genet 23: 271–375 (1985).
12. Feinberg AP, Vogelstein B: A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. Anal Biochem 132: 6–13 (1983).
13. Fryxell PA: A revised taxonomic interpretation of *Gossypium* L. (Malvaceae). Rheedea 2: 108–165 (1992).
14. Fryxell PA: The natural history of the cotton tribe. Texas A&M University Press, College Station, TX (1979).
15. Ganal MW, Broun P, Tanksley SD: Genetic mapping of tandemly repeated telomeric DNA sequences in tomato (*Lycopersicon esculentum*). Genomics 14: 444–448 (1992).

16. Hall KJ, Parker JS, Ellis THN: The relationship between genetic and cytogenetic map of pea. I. Standard and translocation karyotypes. Genome 40: 744–754 (1997).
17. Hall KJ, Parker JS, Ellis THN, Turner T, Knox MR, Hofer JMI, Lu J, Ferrandiz C, Hunter PJ, Taylor JD, Baird K: The relationship between genetic and cytogenetic map of pea. II. Physical maps of linkage mapping populations. Genome 40: 755–769 (1997).
18. Hanson RE, Islam-Faridi MN, Percival EA, Crane CF, Ji Y, McKnight TD, Stelly DM, Price HJ: Distributions of 5S and 18S-28S rDNA loci in a tetraploid cotton (*Gossypium hirsutum* L.) and its putative diploid ancestors. Chromosoma 105: 55-61 (1996).
19. Heslop-Harrison JS, Harrison GE, Leitch IJ: Reprobing of DNA:DNA *in situ* hybridization preparations. Technical Tips 8: 372–373 (1992).
20. Islam-Faridi MN, Mujeeb-Kazi A: Visualization of *Secale cereale* DNA in wheat germplasm by fluorescent *in situ* hybridization. Theor Appl Genet 90: 595–600 (1995).
21. Jewell DC, Islam-Faridi MN: Details of a technique for somatic chromosome preparation and C-banding of maize. In: Freeling M, Walbot W (eds) The Maize Handbook, pp. 484–493. Springer-Verlag, New York (1994).
22. Kimber G: Basis of the diploid-like meiotic behavior of polyploid cotton. Nature 191: 98–99 (1961).
23. Kohler MR, Vogt PH: Interstitial deletions of repetitive DNA blocks in dicentric human Y chromosomes. Chromosoma 103: 324–330 (1994).
24. Nickrent DL, Soltis DE: A comparision of angiosperm phylogenies from nuclear 18S rRNA and rbcL sequences. Ann MO Bot Gard 82: 208–234 (1995).
25. Nishioka Y, Dolan BM, Zahed L: Molecular characterization of a mouse Y chromosomal repetitive sequence amplified in distantly related species in the genus *Mus*. Genome 36: 588–593 (1993).
26. Nishioka Y, Lamothe E: Isolation and characterization of a mouse Y chromosomal repetitive sequence. Genetics 113: 417–432 (1986).
27. Nishioka Y, Lamothe E: Isolation of human Y chromosomal major repetitive sequences from a flow-sorted Y chromosomal library. Am J Med Genet 27: 711–717 (1987).
28. Paterson AH, Brubaker CL, Wendel JF: A rapid method for extraction of cotton (*Gossypium* spp.) genomic DNA suitable for RFLP or PCR analysis. Plant Mol Biol Rep 11: 122–127 (1993).
29. Reinisch AJ, Dong J-M, Brubaker CL, Stelly DM, Wendel JF, Paterson AH: A detailed RFLP map of cotton, *Gossypium hirsutum* × *Gossypiumbarbadens*: chromosome organization and evolution in a disomic polyploid genome. Genetics 138: 829–847 (1994).
30. Roder MS, Lapitan NL, Sorrells ME, Tanksley SD: Genetic and physical mapping of barley telomeres. Mol Gen Genet 238: 294–303 (1993).
31. Royo J, Nass N, Matton DP, Okamoto S, Clarke AE, Newbigin E: A retrotransposon-like sequence linked to the *S*-locus of *Nicotiana alata* is expressed in styles in response to touch. Mol Gen Genet 250: 180–188 (1996).
32. Sanger F, Nicklen S, Coulson AR: DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci USA 74: 5363–5367 (1977).
33. Stallings RL, Whitmore SA, Doggett NA, Callen DF: Refined physical mapping of chromosome 16-specific low-abundance repetitive DNA sequences. Cytogenet Cell Genet 63: 97–101 (1993).
34. Wendel JF: New World cottons contain Old World cytoplasm. Proc Natl Acad Sci USA 86: 4132–4136 (1989).
35. Wendel JF, Albert VA: Phylogenetis of the cotton genus (*Gossypium*): character-state weighted parsimony analysis of chloroplast-DNA restriction site data and its systematic and biogeographic implications. Syst Bot 17: 115–143 (1992).
36. Wendel JF, Schnabel A, Seelanan T: Bi-directional interlocus concerted evolution following allopolyploid speciation in cotton (*Gossypium*). Proc Natl Acad Sci USA 92: 280–284 (1995).
37. Wendel JF, Schnabel A, Seelanan T: An unusual ribosomal DNA sequence from *Gossypium gossypioides* reveals ancient, cryptic, intergenomic introgression. Mol Phyl Evol 4: 298–313 (1995).
38. Wu KS, Tanksley SD: Genetic and physical mapping of telomeres and macrosatellites of rice. Plant Mol Biol 22: 861–872 (1993).
39. Zhao X, Wing RA, Paterson AH: Cloning and characterization of the majority of repetitive DNA in cotton (*Gossypium* L.). Genome 38: 1177–1188 (1995).
40. Zhao X, Wu T, Xie Y, Wu R: Genome-specific repetitive sequences in the genus *Oryza*. Theor Appl Genet 78: 201–209 (1989).
41. Zhao X, Zhang H-B, Wing RA, Paterson AH: A simple method for isolation of megabase DNA from cotton. Plant Mol Biol Rep 12: 110–115 (1994).
42. Zhao X, Si Y, Hansen RE, Crane CF, Rice HJ, Stelly DM, Wendel JF, Paterson AH: Dispersed repetative DNA has spread to new genomes in cotton. Genome Res 8: 479–492 (1998).