Theor Appl Genet (1998) 96: 832–839

J. T. Miller · S. A. Jackson · S. Nasuda
B. S. Gill · R. A. Wing · J. Jiang

Cloning and characterization of a centromere-specific repetitive DNA element from *Sorghum bicolor*

Received: 11 November 1997 / Accepted: 17 November 1997

Abstract A 823-bp *Sau*3AI fragment (pSau3A10) was subcloned from a sorghum bacterial artificial chromosome (BAC) clone, 13I16, that contains DNA sequences specific to the centromeres of grass species. Sequence analysis showed that pSau3A10 consists of six copies of an approximately 137-bp monomer. The six monomers were organized into three dimers. The monomers within the dimers shared 62–72% homology and the dimers were 79–82% homologous with each other. Fluorescence in situ hybridization (FISH) analysis indicated that the *Sau*3A10 family is present only in the centromeres of sorghum chromosomes. Sequencing, Southern hybridization, and Fiber-FISH analyses indicated that the *Sau*3A10 family is tandemly arranged and is present in uninterrupted stretches of up to at least 81 kb of DNA. Slot-blot analysis estimated that the Sau3A10 family constitutes 1.6–1.9% of the sorghum genome. The long stretches of *Sau*3A10 sequences were interrupted by other centromeric DNA elements. Southern analysis indicated that the *Sau*3A10 sequence is one of the most abundant DNA families located in sorghum centromeres and is conserved only in closely related sorghum species. Methylation experiments indicated that the cytosine of the CG sites in sorghum centromeric regions is generally methylated. The structure and organization of the *Sau*3A10 family shared similarities with centromeric DNA repeats in other eukaryotic species. It is suggested that the *Sau*3A10 family is probably an important part of sorghum centromeres.

Key words *Sorghum bicolor* · Repetitive DNA sequences · Centromeres · Molecular cytogenetics

Communicated by G. Wenzel

J. T. Miller · S. A. Jackson · J. Jiang (✉)
Department of Horticulture, University of Wisconsin-Madison, Madison, WI 53706, USA

S. Nasuda · B. S. Gill
Department of Plant Pathology, Kansas State University Manhattan, KS 66506, USA

R. A. Wing
Department of Agronomy and Biological Sciences, Clemson University, Clemson, SC 29634, USA

Introduction

The centromere is a specialized subdomain of the eukaryotic chromosomes. The principle function of centromeres is to ensure proper chromosome segregation during cell division by serving as the site for kinetochore assembly and sister chromatid attachment. Organization of centromeric DNA has been analyzed in detail in yeast, *Drosophila melanogastor*, and humans. The centromeres of budding yeast (*Saccharomyces cerevisiae*) chromosomes consist of 125 base pairs (bp) of DNA without any repetitive elements (Clarke 1990). In contrast, the centromeres of the fission yeast (*Schizosaccharomyces pombe*) chromosomes include several classes of repetitive sequences spanning 40–100 kilobases (kb) of DNA (Clarke et al. 1986; Nakaseko et al. 1986, 1987; Fishel et al. 1988; Chikashige et al. 1989). Single-copy and repetitive DNA sequences specific to each centromere were also detected (Clarke and Baum 1990). The repetitive elements in fission yeast centromeres are arranged in chromosome-specific patterns and the repetitive element K was found to be critical for centromere function (Baum et al. 1994).

Molecular analysis of minichromosomes from *D. melanogastor* revealed that the essential core of the centromere is contained within a 220-kb region (Murphy and Karpen 1995). This essential core, called the *Bora Bora* island, contains significant amounts of complex DNA, consisting of single-copy and middle-repetitive sequences (Le et al. 1995). Normal chromosome stability also requires about 200 kb of DNA on

either side of the essential core. The flanking DNA predominantly contains the AATAT satellite family in *Drosophila*.

Human centromeres predominantly contain alpha satellite (alphoid) DNA. The approximately 170-bp basic unit of the alphoid DNA is tandemly arranged and diverged into chromosome-specific patterns (Willard and Waye 1987; Wevrick and Willard 1989; Oakey and Tyler-Smith 1990). Some centromeres contain up to 4 megabases (Mb) of alphoid DNA. Harrington et al. (1997) constructed fully functional human artificial chromosomes by combining long alphoid DNA arrays with telomeric and genomic DNA. This result showed that a long stretch of alphoid DNA alone can act as a functional human centromere. Alphoid DNA contains a 9-bp motif (B-box) that binds human centromere protein B (CENP-B) in vitro. This motif is present in centromeric DNA from divergent mammalian species (Kipling 1997).

Jiang et al. (1996 b) reported a repetitive DNA element conserved in the centromeres of grass species. This element is a 745-bp fragment (pSau3A9) that was subcloned from a bacterial artificial chromosome (BAC) clone, 13I16 (re-named from 52A4), of sorghum, *Sorghum bicolor*. Its repetitive nature and its conservation in distantly related plant species suggest that the *Sau*3A9 family may play a role in centromere function. Aragon-Alcaide et al. (1996) reported a similar repetitive DNA element that is located in the centromeric regions of cereal chromosomes. Here we report the cloning and characterization of a second repetitive DNA element, pSau3A10, from sorghum centromeres.

Materials and methods

Plant materials

The genus *Sorghum* is generally divided into five sections (Garber 1954). Cultivated sorghum [*S. bicolor* (2n = 20)] is included in the section *Sorghum*. Sorghum cultivar 607E was used in this study. Wild and weedy species of section *Sorghum* analyzed in this study include *Sorghum sudanense* (2n = 20), *Sorghum halepense* (Johnsongrass) (2n = 40), *Sorghum propinquum* (2n = 20), and *Sorghum verticilliflorum* (2n = 20). *Sorghum versicolor* (2n = 10), from the section *Parasorghum*, and *Zea mays* (2n = 20, line B73) were also included.

Plasmid cloning and sequencing

DNA fragments were excised from agarose gels and purified using a Geneclean II kit (Bio 101). The DNA fragments then were ligated to linear pUC18 plasmids and transformed into *Escherichia coli* strain DH5α. White, recombinant, clones on 5-bromo-4-chloro-3-indolyl β-D-galactoside/isopropyl β-D-thiogalactoside plates were analyzed for insert presence by plasmid minipreparation, restriction enzyme digestion and agarose-gel electrophoresis.

Double-stranded plasmid DNA was prepared for sequencing with a QIAprep spin miniprep kit (Qiagen Inc.). Cycle sequencing reactions were performed using Applied Biosystems Ampli*Taq* DNA polymerase, FS Dye Terminator Ready Reactions kit and a Perkin-Elmer Thermocycler (model 2400). Reaction products were analyzed on an Applied Biosystems DNA sequencer (model 373).

Southern blotting

Genomic DNA was isolated from young leaf tissue following protocols described by Giannatasio and Spooner (1994), except that 6× was substituted for 2× CTAB solution. DNA was digested with restriction enzymes according to the manufacturer's instructions and transferred to a Gene-clean nef-988 membrane. Pre-hybridization and hybridization were performed at 65°C in 5× SSC, 0.5% SDS, 0.02 M $NaPO_4$ (pH 6.5), 2 mM EDTA, 10 mM Tris pH 7.4, and 0.02% denatured salmon-sperm DNA. Probes were labeled with ^{32}P and hybridized overnight at 65°C. All washes were performed at 65°C in 0.1% SDS with the stringency controlled by varying concentrations of SSC (from 2× to 0.1×). Membranes were exposed to X-ray film.

Slot-blot hybridization

Copy numbers of the *Sau*3A10 repeat in the sorghum genome and within BAC clone 13I16 were determined by slot-blot hybridization. Standard dilutions of the pSau3A10 probe were compared to known quantities of genomic DNA and BAC DNA. All DNA samples were quantified with a spectrophotometer. DNA denaturation and neutralization followed Zhao et al. (1989). DNA was transferred through a slot-blot template onto a Gene-clean nef-988 membrane. Pre-hybridization and hybridization followed the procedure outlined in the previous section. Filters were washed for 15 min each in 2× SSC, 0.1% SDS; 0.5× SSC, 0.1% SDS; and 0.2× SSC, 0.1% SDS at 65°C and then exposed to X-ray film for 24 h. Band intensities were measured on the autoradiographs by IPLab Spectrum v3.1 software.

Fluorescence in situ hybridization (FISH)

Detailed protocols for chromosome preparation, probe labeling, in situ hybridization, and signal detection were described previously (Jiang et al. 1996 a). Isolation of sorghum nuclei and preparation of extended DNA fibers were according to Liu and Whittier (1994) and Fransz et al. (1996), respectively. Images were captured with a Photometrics SenSys CCD (charge coupled device) camera coupled to a Macintosh computer. IPLab Spectrum v3.1 software was used to capture fluorescent signals and to measure the size of FISH signals on extended DNA fibers.

Results

Cloning and sequencing of the *Sau*3A10 repetitive DNA family

DNA of BAC clone 13I16 was digested with several restriction enzymes and blotted to a nylon membrane. A *Sau*3A I fragment strongly hybridized to sorghum genomic DNA was subcloned and named pSau3A10.

Sequencing analysis showed that pSau3A10 contains an 823-bp insert consisting of three copies of a tandemly arranged repeat with a consensus sequence of 277 bp (Fig. 1A). The sequence has a 42% GC content. The three tandemly arranged members are 79–82% homologous with each other (Table 1). The entire clone is

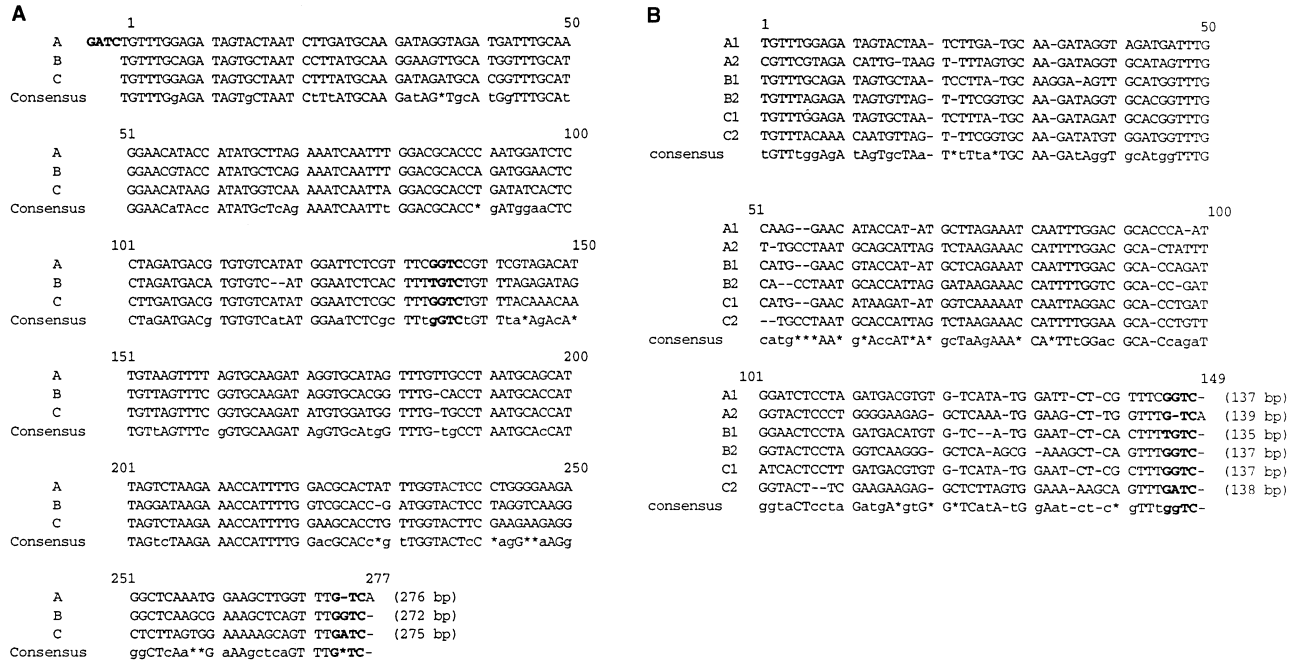

Fig. 1A, B Complete sequence of the 823-bp clone pSau3A10. **A** The clone consists of three tandemly arranged dimers (A–C) of 276, 272, and 275 bp, respectively, with a consensus sequence of 277 bp. In the consensus sequence, *lower case* represents the majority nucleotide and *upper case* indicates that nucleotide found in all sequences. A * indicates that there is no consensus nucleotide. About 72% of the sites are conserved among the three dimers. Each dimer is defined by the mutated *Sau*3AI (GATC) sites marked in *bold* and located at the end of the dimers (273–276). The monomers are defined by the mutated *Sau*3AI sites marked in bold and located in the middle (134–137) of the dimers. A *Hin*dIII restriction site (AAGCTT) is found in dimer A (262–267). **B** The six monomers, A1–C2, have a consensus sequence of 149 bp. The order of the six monomers in pSau3A10 from 5′-3′ is A1, A2, B1, B2, C1, C2; 14 of the 15 sites marked by a * show conservation at the dimer level but not at the monomer level

Table 1 Comparisons of homology between different repeats within the 823-bp clone pSau3A10. **A** Homology of the 277-bp dimer sequences given as a percentage of the aligned sequences. **B** Homology of the 137-bp monomer sequences given as a percentage of the aligned sequences

A

Sequence	B	C
A	79%	81%
B	–	82%

B

Sequence	A2	B1	B2	C1	C2
A1	64%	83%	70%	83%	63%
A2		66%	76%	66%	76%
B1			72%	84%	65%
B2				69%	79%
C1					62%

flanked by *Sau*3A I (GATC) restriction sites. The internal sites correlating to *Sau*3A I differ only at the second base (A) (Fig. 1A). Areas of high homology (base pairs 71–89 and 206–221 of the consensus sequence) are spread throughout the sequence and are separated by areas of less homology (32–43 and 239–268).

The 277-bp repeat is structurally a dimer consisting of two tandemly arranged members of a basic monomeric repeat. The length of the six sequenced monomers varied from 135 to 139 bp. The consensus sequence of the monomers is 10 bp longer (149 bp) than the longest monomer (139 bp), due to numerous small insertion/deletion events. The *Sau*3AI sites defining the monomers have conserved TC dinucleotides; five of six of these have G at the first base and four of six have G at the second base, indicating that the ancestral GGTC mutated into the *Sau*3AI site (Fig. 1 B). The homology among monomers within the dimers is from 62% to 72% (Table 1) and is less than among dimers. For example, the first half of the first dimer, A1, has a 64% similarity to A2 but a 83% homology to B1. There are 14 nucleotide sites that are conserved in the dimers but not in the monomers (Fig. 1 B). These results suggest that the dimer is the primary unit of amplification. Searches of the GenBank database found no significant matches with the pSau3A10 sequence.

Organization of the *Sau*3A10 family

Southern-blot analysis of sorghum genomic DNA digested with the *Hin*dIII restriction enzyme and probed

with pSau3A10 showed ladder-like patterns which are characteristic for tandem repeats (Figs. 2 and 4). A *Hin*dIII restriction site (AAGCTT) is found at nucleotides 262–267 of the first dimer (Fig. 1 A). Southern hybridization showed multiples of the 277-bp repeat as dimer, tetramer and higher-level tandem arrays in the *Hin*dIII digest (Fig. 2 and 4). Monomer (137 bp) and trimer (411 bp) bands were not found. This result indicates that the dimer, rather than the monomer, is the basic unit of amplification of this DNA family in the sorghum genome.

FISH results showed that pSau3A10 hybridized exclusively to the centromeric regions of all sorghum chromosomes (Fig. 3A–C), indicating that the *Sau*3A10 family is specific to the centromeric regions.

Fiber-FISH analysis revealed that the *Sau*3A10 sequence is organized into blocks of different sizes (Fig. 3 D). The largest observed block, with small interspersed gaps (< 2 μm) was 27.36-μm long. Using known *Arabidopsis thaliana* cosmid clones, the resolution of Fiber-FISH was calibrated in our laboratory as 2.96 kb/μm, and gaps that are smaller than 2 μm are considered as continuous DNA sequences (our unpublished data). The resolution of Fiber FISH in *A. thaliana* suggests that a 27.36–μm-long block represents about 81 kb of DNA and contains approximately 590 monomers of the *Sau*3A10 family. Fiber-FISH signals as long as 350 μm, corresponding to about 1 Mb of DNA, with various sizes of gaps were observed, indicating that the *Sau*3A10 family may be distributed throughout the sorghum centromeric regions.

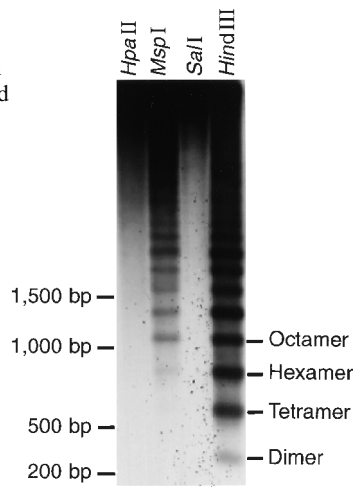

Fig. 2 Sorghum genomic DNA was digested with three methylation-sensitive restriction enzymes (*Hpa*II, *Msp*I, *Sal*I) and one methylation-insensitive enzyme (*Hin*dIII) and the blot was hybridized with pSau3A10

Specificity of the *Sau*3A10 family

The pSau3A10 probe hybridized only to the sorghum genomic DNA and not to the DNA from several other grass species, including, rice, sugarcane, bamboo, rye and wheat (data not shown). A number of wild species in the genus *Sorghum* were analyzed for the presence of DNA families homologous to the *Sau*3A10 family. pSau3A10 hybridized strongly to five species in the section *Sorghum*, including *S. bicolor*, *S. sudanense*, *S. verticilliflorum*, *S. halepense* and *S. propinquum*. However, pSau3A10 did not hybridize to genomic DNA from *S. versicolor* (section *Parasorghum*) or to maize DNA (Fig. 4). All the *Sorghum* species analyzed shared similar hybridization patterns.

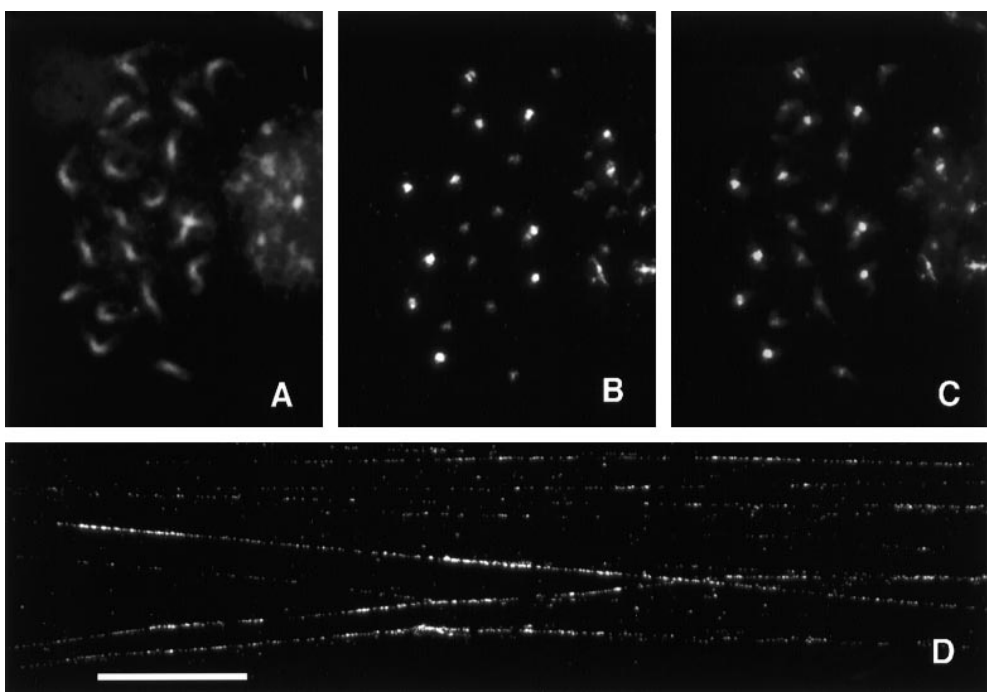

Fig. 3A–C FISH analysis of sorghum prometaphase chromosomes using pSau3A10 as a probe. **A** prometaphase chromosomes; **B** FISH signals; **C** a merged image from A and B. Ten chromosomes show stronger centromeric signals than the other ten chromosomes. **D** Fiber-FISH analysis using pSau3A10 as a probe. The *Sau*3A10 family is organized into blocks of different sizes. Bar represents 20 micrometers. Gaps smaller than 2 micrometers may represent continuous sequences

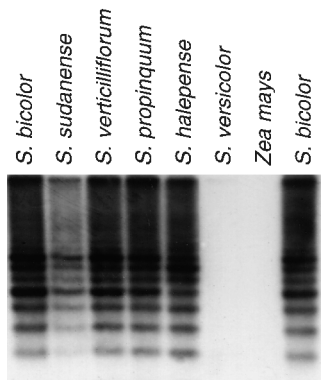

Fig. 4 DNA from sorghum, maize, and five wild *Sorghum* species was digested with *Hin*dIII and the blot was hybridized with pSau3A10

FISH analysis in *S. bicolor* showed that 10 of the 20 sorghum chromosomes had strong signals at the centromeres. The signals on the other ten centromeres were relatively weak (Fig. 3B). There are two possible explanations for these results: (1) half of the sorghum centromeres contain more copies of the *Sau*3A10 family than the other half, or else (2) ten sorghum chromosomes have a subset of the *Sau*3A10 family which is only partially homologous to pSau3A10; this partial homology then resulted in the relatively weak FISH signals.

Abundance of the *Sau*3A10 family in the sorghum genome

Southern blotting of BAC 13I16 DNA digested with *Hae*II and *Hin*dIII and probed with pSau3A10 showed that about half of the fragments, excluding those derived from the BAC vector, hybridized to pSau3A10 (Fig. 5). When the same blot was probed with sorghum genomic DNA, the hybridization patterns were almost identical to those probed by pSau3A10 (Fig. 5). About ten distinct DNA elements have been identified in BAC 13I16. The amount of DNA from other centromeric families in the sorghum genomic DNA probe (about 50 ng) are possibly negligible as compared to the DNA from the Sau3A10 family (1.6–1.9% of the sorghum genome, see below). Thus, only fragments containing the *Sau*3A10 sequence were lit up when sorghum genomic DNA was used as a probe. These results suggest that Sau3A10 is possibly one of the most abundant DNA families located in sorghum centromeres.

The copy numbers of the *Sau*3A10 family in the BAC clone 13I16 and in the sorghum genome were estimated by slot-blot analysis (Fig. 6). The copy number was determined by comparing the band intensities of known amounts of pSau3A10 to known amounts of genomic DNA and BAC 13I16 DNA. The hybridization intensity of genomic DNA equivalent to 2000 sorghum genomes was close to that from 2×10^8 monomers of the *Sau*3A10 family, suggesting 88 000 monomers per haploid genome. Based on a sorghum genome size of 772 Mb (Arumganathan and Earle 1991), the

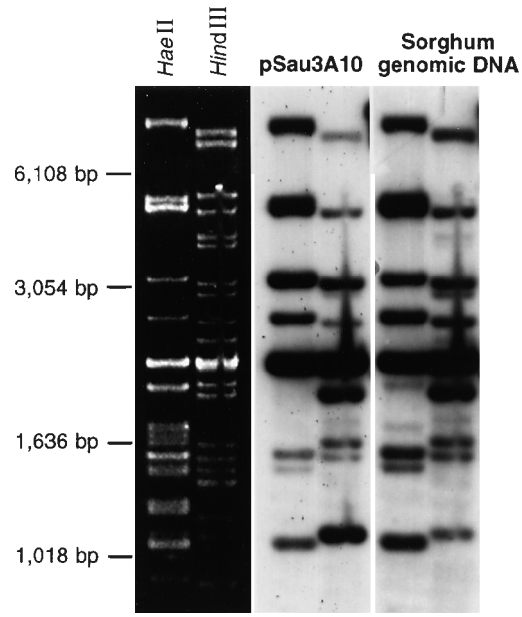

Fig. 5 DNA from sorghum BAC clone 13I16 was digested with *Hae*II and *Hin*dIII (left). The same blot was probed with pSau3A10 (center) and sorghum genomic DNA (right)

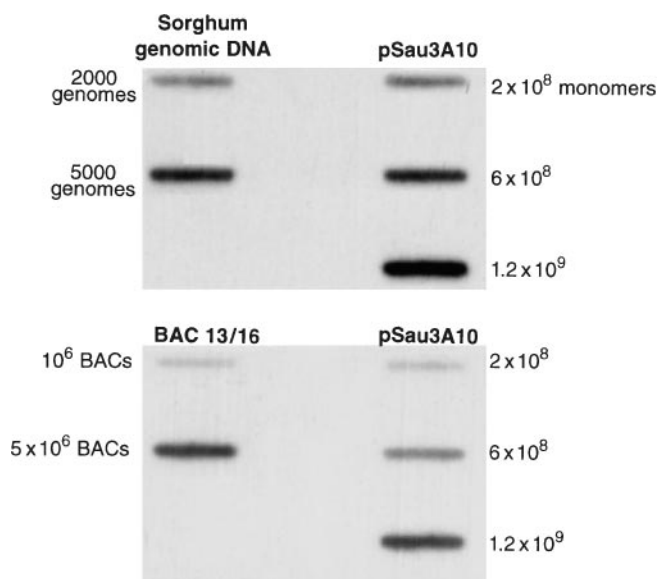

Fig. 6 Slot-blot analysis. Copy numbers of the monomer of the *Sau*3A10 family in the sorghum genome and BAC 13I16 were determined by the comparison of known quantities of sorghum genomic DNA (top) and BAC 13I16 DNA (bottom) to a serial dilution of known quantities of the 137-bp monomer using probe pSau3A10

*Sau*3A10 family comprises 1.6% of the sorghum genome. A second comparison found that 5000 sorghum genome-equivalents was close to 6×10^8 monomers, suggesting 109 000 copies/haploid or 1.9% of the sorghum genome. Therefore the *Sau*3A10 family

comprises between 1.6 and 1.9% of the sorghum genome. BAC 13I16 was estimated to contain 240 copies of the monomer because 1.2×10^7 copies of the BAC DNA has a similar hybridization intensity to 6×10^8 copies of the *Sau*3A10 monomer. This corresponds to 33 kb of clone 13I16 which is approximately 90 kb long.

A sorghum BAC library (Woo et al. 1994) was screened with probes pSau3A9 and pSau3A10. About 5.4% of the BAC clones in the library (13 440 clones in total) hybridized to pSau3A10, while only 0.6% of the clones hybridized to pSau3A9.

Methylation of the centromeric regions in sorghum

The cytosine nucleotide is the most common site for methylation in plant genomes. The most common dinucleotide site for cytosine methylation is CG, but methylation occurs at lower frequencies when the C and G are separated by one to two A or T bases (Gruenbaum et al. 1981). Figure 2 shows different levels of cytosine methylation in the centromeric regions of *Sorghum*. *Msp*I and *Hpa*II are isoschizomer restriction enzymes that recognize the 5′CCGG3′ sequence. Neither can cut when the 5′C is methylated. Only *Msp*I, not *Hpa*II, can cleave when the internal cytosine is methylated. When *Hpa*II- and *Msp*I digested sorghum DNA was transferred to a nylon membrane and probed with the centromeric DNA probe pSau3A10, a relatively faint ladder as compared to the *Hin*dIII lane was found in the *Msp*I lane but not in the *Hpa*II-lane (Fig. 2). This result indicates that 5′CCGG3′ sites are present in some of the Sau3A10 monomers even though they are not found in the probe pSau3A10. Detection of bands in the *Msp*I lane, but not in the *Hpa*II lane, suggests that the internal cytosine of the 5′CCGG3′ sequence is generally methylated. The restriction enzyme *Sal*I contains the CG motif (GTCGAC) and fragments hybridized to pSau3A10 were not detected in the *Sal*I lane (Fig. 2).

Discussion

Highly repetitive tandem repeats specific to centromeric regions have been identified in both mammalian and plant species. Human alphoid DNA is a tandemly repeated family with a basic 170-bp unit comprising 3% of the human genome. Alphoid DNA can exist in uninterrupted stretches of thousands of copies of the 170-bp repeat (Warburton and Willard 1990). Tandem groups of the basic repeat can sometimes become an amplified unit. For example, a dimer consisting of two tandem 170-bp human-alphoid DNA monomers has been identified due to a mutation in the *Eco*RI site that separates the monomers. This dimer is a major unit of amplification and 60% of the alphoid DNA exists as this dimer. The dimers share 80–87% homology with each other, while adjacent monomers within a dimer share 55–70% homology (Shmookler Reuse et al. 1985). Higher-level units, such as tetramers defined by the same *Eco*RI site, comprise 30% of alphoid DNA. Chromosomal areas can have unique arrangements of higher- and lower-level repeat units (Warburton and Willard 1990). Old world and new world primates are thought to have diverged 45–50 million years ago. Their alphoid DNA maintains an overall 64% sequence identity with small highly conserved areas within the repeat. The predominance of dimers and tetramers is also maintained in new world primates (Alves et al. 1994).

A 180-bp tandem repeat family, pAL1, isolated from *A. thaliana* (Martinez-Zapater et al. 1986), is the most abundant repetitive DNA family in this species. The pAL1 family has about 5000 copies comprising 1–1.5% of the *A. thaliana* genome. Homology among monomers range from 87% to 92% (Martinez-Zapater et al. 1986). Estimations of the length of the uninterrupted pAL1 repeat ranges from 50 kb (Martinez-Zapater et al. 1986) to 1200 kb (Murata et al. 1994). FISH analysis showed that the pAL1 family is located at the paracentromeric regions of all five *A. thaliana* chromosomes (Maluszynska and Heslop-Harrison 1991). Unlike human alphoid DNA, the pAL1 family is interrupted by other repetitive sequences (Brandes et al. 1997). Southern hybridization of pAL1 to *Arabidopsis arenosa* DNA showed a very weak ladder as compared to *A. thaliana*, reflecting the high divergence of this DNA family within these closely related species (Kamm et al. 1995). A centromeric tandem repeat (pAa214) was isolated from *A. arenosa*. This repeat maintains 58% sequence identity to the pAL1 family (Kamm et al. 1995).

The *Sau*3A10 family identified in this study shared many characteristics with the human alphoid DNA and the *A. thaliana* pAL1 family. The *Sau*3A10 family is also tandemly arranged in large arrays. As in the pAL1 family, the *Sau*3A10 sequence is disrupted by other sorghum centromeric DNA elements, including the *Sau*3A9 family (unpublished data). The *Sau*3A10 sequence is present only in closely related sorghum species, indicating a fast divergence of this repeat. The monomer (137 bp) of the *Sau*3A10 family is shorter than monomers of the *Arabidopsis* pAL1 repeat (180 bp) and human alphoid satellite DNA (170 bp). Like the human alphoid DNA, the majority of the *Sau*3A10 family is organized as dimers in the sorghum genome. The homology among *Sau*3A10 monomers within the dimers (62–72%) is similar to adjacent monomers of alphoid DNA (55–70%). Likewise, homology among *Sau*3A10 dimers (79–82%) is similar to adjacent dimers of alphoid DNA (80–87%; Shmookler Reuse et al. 1985).

Centromeric regions usually consist of heterochromatin that is thought to be highly methylated. Human alphoid centromeric arrays are highly methylated at the CG dinucleotide as shown by the lack of

hybridizing bands in methylation-sensitive *Hpa*II-digested DNA when compared to *Msp*I-digested DNA (Shmookler Reuse et al. 1985). Cytosine methylation is also common in plants, with the CG motif most commonly methylated (Gruenbaum et al. 1981). The *A. thaliana* genome has only about 15% repetitive DNA sequences, the lowest known in plants. This low amount of repetitive sequences correlates to low rates of methylation, as shown by digestion profiles of enzymes with differential methylation sensitivity. Leutwiler et al. (1984) determined that only 4.6% of *A. thaliana* cytosines are methylated compared to 20.1% in wheat germ (Gruenbaum et al. 1981). However, when the digested *A. thaliana* DNA was blotted and probed with the pAL1 repeat, only *Msp*I, and not *Hpa*II, had the characteristic ladder of a repeated sequence (Martinez-Zapater et al. 1986). This indicated that methylation of the internal cytosine of the *Msp*I/*Hpa*II restriction site, CCGG, is a prominent feature of the *A. thaliana* centromeres.

The methylation status of sorghum centromeric regions has been analyzed in the present study. Although CCGG sites are not present in the p*Sau*3A10 sequence, a faint ladder pattern derived from *Msp*I-digested sorghum DNA indicated that CCGG sites are present in some of the *Sau*3A10 monomers. Comparison of Southern hybridization patterns from *Msp*I-, *Hpa*II-, and *Sal*I-digested sorghum DNA suggested that the cytosine of CG sites in sorghum centromeres is generally methylated. The DNA fragments hybridized to p*Sau*3A10 in the *Msp*I-digested sorghum DNA were faint and bands corresponding to dimers and tetramers were not clearly visible (Fig. 2). At present we cannot determine whether this is the result of partial methylation of the external cytosine of the CCGG sites in sorghum centromeres or lack of pairs of CCGG sites flanking dimers or tetramers.

The function of cloned centromeric DNA can be tested using the elegant genetic systems of budding and fission yeast (Clarke and Carbon 1980; Clarke and Baum 1990). However, functional analysis of cloned centromeric DNA is difficult in other eukaryotic species. Because of the lack of convincing functional results, the centromeric role of the human alphoid DNA was debated for many years until the construction of human artificial chromosomes (Harrington et al. 1997). Although there is no evidence to indicate that the pAL1 family is part of *A. thaliana* centromeres, its nature as the most dominant DNA component in the centromeric regions suggests that it may play a role in centromere function. The organization of the *Sau*3A10 family in sorghum shares a number of similarities with human alphoid DNA and the pAL1 family in *A. thaliana*. The present study also indicates that the *Sau*3A10 family is probably the most abundant DNA family in sorghum centromeres. These results suggest that the *Sau*3A10 family is probably an important part of the sorghum centromeres.

Acknowledgements We thank Dr. G. H. Liang for providing seeds of the *Sorghum* species. We are grateful to Dr. M. J. Havey and Dr. S. M. Kaeppler for critical reading of this paper. The research is supported by Funds 135-0534 and 135-0528 from the Graduate School of the University of Wisconsin-Madison.

References

Alves G, Seuánez HN, Fanning T (1994) Alpha satellite DNA in neotropical primates (*Platyrrhini*). Chromosoma 103: 262–267

Aragon-Alcaide L, Miller T, Schwarzacher T, Reader S, Graham M (1996) A cereal centromeric sequence. Chromosoma 105: 261–268

Arumuganathan K, Earle ED (1991) Nuclear DNA content of some important plant species. Plant Mol Biol Rep 9: 208–219

Baum M, Ngan VK, Clarke L (1994) The centromeric K-type repeat and the central core are together sufficient to establish a functional *Schizosaccharmyces pombe* centromere. Mol Biol Cell 5: 747–761

Brandes A, Thompson H, Dean C, Heslop-Harrison JS (1997) Multiple repetitive DNA sequences in the paracentric regions of *Arabidopsis thaliana* L. Chromosome Res 5: 238–246

Chikashige Y, Kinoshita N, Nakaseko Y, Matsumoto T, Murakami S, Niwa O, Yanagida, M (1989) Composite motifs and repeat symmetry in *S. pombe* centromeres: direct analysis by integration of *Not*I restriction sites. Cell 57: 739–751

Clarke L (1990) Centromeres of budding and fission yeasts. Trends Genet 6: 150–154

Clarke L, Baum M (1990) Functional analysis of a centromere from fission yeast: a role for centromere-specific repeated DNA sequences. Mol Cell Biol 10: 1863–1872

Clarke L, Carbon J (1980) Isolation of a yeast centromere and construction of functional small circular chromosomes. Nature 287: 504–509

Clarke L, Amstutz H, Fishel B, Carbon J (1986) Analysis of centromeric DNA in the fission yeast *Schizosaccharomyces pombe*. Proc Natl Acad Sci USA 83: 8253–8257

Fishel B, Amstutz H, Baum M, Carbon J, Clarke L (1988) Structural organization and functional analysis of centromeric DNA in the fission yeast *Schizosaccharomyces pombe*. Mol Cell Biol 8: 754–763

Fransz PF, Alonso-Blanco C, Liharska TB, Peeters AJM, Zabel P, de Jong JH (1996) High-resolution physical mapping in *Arabidopsis thaliana* and tomato by fluorescence in situ hybridization to extended DNA fibres. Plant J 9: 421–430

Garber ED (1954) Cytotaxonomic studies in the genus *Sorghum*. III. The polyploid species of the subgenera *Parasorghum* and *Stiposorghum*. Bot Gaz 115: 336–342

Giannattasio RB, Spooner DM (1994) A re-examination of species boundaries and hypotheses of hybridization concerning *Solanum megistacrolobum* and *S. toralapanum* (*Solanum* sect. *Petota*, series *Megistacrolobum*): molecular data. Systematic Bot 19: 106–115

Gruenbaum Y, Naveh-Many T, Cedar H, Razin A (1981) Sequence specificity of methylation in higher plant DNA. Nature 292: 860–862

Harrington JJ, Bokkelen GV, Mays RW, Gustashaw K, Willard HF (1997) Formation of *de novo* centromeres and construction of first-generation human artificial microchromosomes. Nature Genet 15: 345–355

Jiang J, Hulbert SH, Gill BS, Ward DC (1996 a) Interphase fluorescence in situ mapping: a physical mapping strategy for plant species with large complex genomes. Mol Gen Genet 252: 497–502

Jiang J, Nasuda S, Dong F, Scherrer CW, Woo S, Wing RA, Gill BS, Ward DC (1996 b) A conserved repetitive DNA element located in the centromeres of cereal chromosomes. Proc Natl Acad Sci USA 93: 14210–14213

Kamm A, Galasso I, Schmidt T, Heslop-Harrison JS (1995) Analysis of a repetitive DNA family from *Arabidopsis arenosa* and relationship between *Arabidopsis* species. Plant Mol Biol 27: 853–862

Kipling D (1997) Centromeres, CENP-B and Tigger too. Trends Genet 13: 141–145

Le M-H, Duricka D, Karpen GH (1995) Islands of complex DNA are widespread in *Drosophila* centric heterochromatin. Genetics 141: 283–303

Leutwiler LS, Hough-Evans BR, Meyerowitz EM (1984) The DNA of *Arabidopsis thaliana*. Mol Gen Genet 194: 15–23

Liu Y, Whittier RF (1994) Rapid preparation of megabase plant DNA from nuclei in agarose plugs and microbeads. Nucleic Acid Res 22: 2168–2169

Maluszynska J, Heslop-Harrison JS (1991) Localization of tandemly repeated DNA sequences in *Arabidopsis thaliana*. Plant J 1: 159–166

Martinez-Zapater JM, Estelle MA, Somerville CR (1986) A highly repeated DNA sequence in *Arabidopsis thaliana*. Mol Gen Genet 204: 417–423

Murata M, Ogura Y, Motoyoshi F (1994) Centromeric repetitive sequences in *Arabidopsis thaliana*. Jpn J Genet 69: 361–370

Murphy TD, Karpen GH (1995) Localization of centromere function in a *Drosophila* minichromosome. Cell 82: 599–609

Nakaseko Y, Adachi Y, Funahashi S, Niwa O, Yanagida M (1986) Chromosome walking shows a highly homologous repetitive sequence present in all the centromere regions of fission yeast. EMBO J 5: 1011–1021

Nakaseko Y, Kinoshita N, Yanagida M (1987) A novel sequence common to the centromere regions of *Schizosaccharomyces pombe* chromosomes. Nucleic Acids Res 15: 4705–4715

Oakey R, Tyler-Smith C (1990) Y-chromosome DNA haplotyping suggests that most European and Asian men are descended from one of two males. Genomics 7: 325–330

Shmookler Reuse RJ, Srivastave A, Beranek DT, Goldstein S (1985) Human alphoid family of tandemly repeated DNA: sequence of cloned tetrameric fragments and analysis of familial divergence. J Mol Biol 186: 31–41

Warburton PE, Willard HF (1990) Genomic analysis of sequence variation on tandemly repeated DNA: evidence for localized homogeneous sequence domains within arrays of α-satellite DNA. J Mol Biol 216: 3–16

Wevrick R, Willard HF (1989) Long-range organization of tandem arrays of a satellite DNA at the centromeres of human chromosomes: high frequency array length polymorphism and meiotic stability. Proc Natl Acad Sci USA 86: 9394–9398

Willard HF, Waye JS (1987) Hierarchical order in chromosome-specific human alpha satellite DNA. Trends Genet 3: 192–198

Woo S-S, Jiang J, Gill BS, Paterson AH, Wing RA (1994) Construction and characterization of a bacterial artificial chromosome library of *Sorghum bicolor*. Nucleic Acids Res 22: 4922–4931

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