FINAL REPORT

Award No. 99-35300-7819
Title: Cytomolecular mapping in sorghum and maize using *in situ* hybridization
PI: Daniel G. Peterson, University of Georgia; Plant Genome Mapping Lab; Athens, GA 30602

Sorghum (*Sorghum bicolor*), the subject of large-scale molecular and physical mapping studies, is a staple crop for millions in the semi-arid tropics and the focus of a 1.5 billion-dollar industry in the U.S. Despite its importance, almost nothing is known about the structure and appearance of sorghum chromosomes. Because the genomes of eukaryotes are always found within the context of chromosomes, an understanding of how genomes function and evolve is inseparably tied to an understanding of the nature of chromosomes themselves.

In our project we initiated research relating the genetic and physical maps of sorghum linkage group C (SLG-C) to their corresponding pachytene (meiotic) chromosome. The primary means to achieve this goal was "cytomolecular mapping", the use of fluorescence *in situ* hybridization (FISH)[1] to accurately visualize the chromosomal locations of single-copy marker sequences (including genes) on hypotonically spread pachytene chromosomes (also known as synaptonemal complexes or SCs). Cytomolecular mapping permits superimposition of a genetic linkage map onto the physical structure of its corresponding chromosome, and it provides correlative information on how specific chromosomal structures influence crossing over. Eventually, the nearly finished physical map of SLG-C (prepared under separate funding) will be superimposed onto the cytomolecular map to create a highly informative "cytophysical map" of the chromosome. A secondary goal of the research was the localization of SLG-C molecular markers on SC spreads of maize with the goal of constructing cytomolecular maps for maize chromosomal regions sharing homology with SLG-C. Information generated in this study and subsequent studies should be of profound interest to plant breeders and geneticists because it will (a) provide insight into the relationships between linkage (recombination) distances, base pair distances, and chromatin structure between molecular markers, (b) provide a means by which the site of a transgene insertion can be rapidly detected and characterized with respect to chromosomal location, (c) help build a more cohesive, comprehensive, and realistic view of eukaryotic genomes, and (d) furnish information about the genome and chromosomes of an agronomically-important species.

RESULTS:

Completed goals: This project significantly advanced knowledge of the genome and meiotic chromosomes of sorghum. A few examples of important discoveries and deliverable research tools include:

(1) Development of a technique for spreading the synaptonemal complexes (SCs ≈ meiotic chromosomes) of sorghum (**Figure 1**). Before this study, SC spreads had never been prepared for sorghum.

(2) Preparation of a sorghum SC karyotype that allows description of individual chromosomes at resolution far surpassing that obtainable using mitotic chromosome squashes (see **Draye et al. 2001**).

(3) Discovery that sorghum linkage group C corresponds to the longest mitotic/meiotic chromosome (#1).

(4) Demonstration of the usefulness of sorghum SC spreads as substrates for fluorescence *in situ* hybridization (FISH) (**Figure 2**).

(5) Completion of the first Cot (DNA renaturation kinetics) analysis of the sorghum genome and isolation of repetitive DNA sequences for use in suppressing repetitive elements in BAC clones used as FISH probes.

(6) Isolation of DNA from roughly 20 BAC clones containing low-copy DNA sequences specific to sorghum linkage group C and labeling of these probes with either biotin or digoxigenin.

Additional accomplishments: *Cot-Based Cloning and Sequencing* (CBCS) is an idea that was developed during construction of the sorghum Cot curve. CBCS, a novel synthesis of Cot analysis (DNA renaturation kinetics), molecular cloning, and high-throughput DNA sequencing, should greatly accelerate comparative genomics research and gene discovery in organisms with repetitive genomes. The steps in CBCS are as follows: (A) a Cot analysis is performed for a species of interest, (B) the results of the Cot analysis are used to guide the hydroxyapatite chromatography-based fractionation of the genome into low-copy and repetitive sequence components, (C) each isolated kinetic component is used to construct a corresponding 'Cot library', and (D) clones from each library are sequenced in numbers proportional to the kinetic (sequence) complexity of the component from which they were derived. For repetitive plant genomes such as maize, pine, cotton, pea, onion, and peanut, CBCS should permit

[1] In FISH, a DNA marker associated with a gene is labeled with a hapten (e.g., biotin, digoxigenin). The labeled marker is then hybridized to chromosomes fixed on glass microscope slides. The chromosomal location of the marker (gene) is visualized using affinity reagents (*e.g.*, antibodies, avidin) labeled with a fluorescent tag(s).

sequencing of all low-copy DNA elements at 4-20 times the efficiency of shotgun sequencing. Unlike methyl-filtration and *Pst*I cloning, CBCS allows low-copy elements to be isolated and sequenced in a manner independent of methylation patterns that vary widely between species, genes within an organism, and developmental stages. In contrast to EST sequencing, CBCS provides access to regulatory sequences and also secures genes independently of their levels or (tissue or organ-specific) patterns of expression, increasing the likelihood of discovering key regulatory genes that are only transiently expressed. In an initial (albeit modest) test of the method, the genome of sorghum (*Sorghum bicolor*) was fractionated into highly repetitive (HR), moderately repetitive (MR), and single/low-copy (SL) sequence components that were consequently cloned to produce HRCot, MRCot, and SLCot genomic libraries. Filter hybridization (blotting) and sequence analysis both show that the HRCot library is enriched in sequences traditionally found in high-copy number (retroelements, rDNA, centromeric repeats), the SLCot library is enriched in low-copy sequences (genes and ESTs), and the MRCot library contains sequences of moderate redundancy (**Figure 3**). Based upon the kinetic complexities of the components, CBCS could be used to capture the sequence complexity of the relatively small sorghum genome (760 Mb) at roughly one-third of the cost of shotgun sequencing. All publications resulting from CBCS research have acknowledged the support of the USDA-NRICGP (**Peterson et al. 2002a, b**). Likewise, a patent application filed by the inventors of CBCS (Daniel G. Peterson, Susan R. Wessler, and Andrew H. Paterson) acknowledges USDA grant 99-35300-7819 as the sole source of extramural funding in the invention's development. CBCS has attracted significant media coverage (see www.msstate.edu/research/mgel/cbcs/press.htm for details).

Complications: Upon starting this project, SC spreads had never been prepared for sorghum. Not surprisingly, preparation of quality SC spreads from sorghum has proven the most difficult, time-consuming, and frustrating aspect of our research. Because SC spreading protocols for maize and other grasses did not work for sorghum, considerable experimentation was required to determine the conditions that would produce sorghum SC spreads. After almost a year of work, we managed to obtain high quality SC spreads (Figure 1). However, we are still working to stablilize/standardize the SC spreading procedure (i.e., a given SC spreading experiment may fail to yield SC spreads although conditions may be 'identical' to those used to generate quality SC spreads the day before). Much to our dismay, we have only been able to obtain SC spreads during the summer months (June, July, & August). Our research indicates that the meiotic substage in which SCs are found becomes so short during non-summer months that it is virtually impossible to obtain SC spreads. Attempts at growing plants in growth chambers mimicking summer conditions have not remedied this problem. Because production of suitable SC spreads is a limiting factor, cytomolecular mapping of SLG-C has not proceeded at the pace we had hoped. Only recently has actual cytomolecular mapping using single-copy probes been possible in sorghum – microscope slides with suitable SC spreads are prepared during the summer and stored at –80ºC for FISH experiments in the non-summer months. A complicating factor is that sorghum SCs typically lack distinguishable kinetochores (centromeres). Consequently, we have had to perform FISH using a centromere-specific probe to position centromeres on the sorghum SCs (**Figure 2**).

Continuing research:
(1) *Producing a SLG-C cytomolecular map* – At present, FISH is being used to position single-copy sequences on SLG-C. The resulting cytomolecular map will be related to the existing sorghum linkage and physical maps. Any publications resulting from current and future SLG-C cytomolecular mapping will cite USDA-NRI award 99-35300-7819.
(2) *Relating sorghum linkage groups to their corresponding pachytene chromosomes* – At present, there are partial physical maps for all ten sorghum linkage groups. Consequently, BAC clones associated with a particular sorghum linkage group presumably could be pooled to prepare a "chromosome-specific FISH paint". FISH paints for each of the ten sorghum linkage groups then could be used to affiliate each linkage group with a cytologically characterized sorghum pachytene chromosome (SC). After getting at least a few cytomolecular markers mapped onto SLG-C, we plan to use chromosome painting to match linkage groups with corresponding SCs.
(3) *Comparison of the cytomolecular map of SLG-C with Zea mays SCs probed with SLG-C BACs* – The construction of partial cytomolecular maps for *Zea mays* using SLG-C BACs as probes was completely dependent upon the successful completion of the SLG-C cytomolecular map. Due to the difficulties associated with spreading sorghum SCs, it is unlikely that research on *Zea mays* will be initiated until cytomolecular mapping in sorghum becomes fairly routine.

PUBLICATIONS RESULTING FROM AWARD 99-35300-7819

Peterson DG, Wessler SR, Paterson AH (2002a) Efficient capture of unique sequences from eukaryotic genomes. *Trends In Genetics*, in press.

Peterson DG, Schulze SR, Sciara EB, Lee SA, Nagel A, Jiang N, Tibbitts DC, Wessler SR, Paterson AH (2002b) Integration of Cot analysis, DNA cloning, and high-throughput sequencing facilitates genome characterization and gene discovery. *Genome Research* **12**: 795-807.

Draye X, Lin Y-R, Qian X-y, Bowers JE, Burow GB, Morrell PL, Peterson DG, Presting GG, Ren S-x, Wing RA, Paterson AH (2001) Toward integration of comparative genetic, physical, diversity, and cytomolecular maps for grasses, using the *Sorghum* genome as a foundation. *Plant Physiology* **125**: 1325-1341.

*Tomkins JP, Peterson DG, Yang TJ, Main D, Ablett EF, Henry RJ, Lee LS, Holton TA, Waters D, Wing RA (2001) Grape (*Vitis vinifera* L.) BAC library construction, preliminary STC analysis, and identification of clones associated with flavonoid and stilbene biosynthesis. *American Journal of Enology and Viticulture* **52**: 287-291.

*Tomkins JP, Peterson DG, Yang T, Main D, Wilkins TA, Paterson AH, Wing RA (2001) Development of genomic resources for cotton (*Gossypium hirsutum*): BAC library construction, preliminary STC analysis, and identification of clones associated with fiber development. *Molecular Breeding* **8**: 255-261.

Peterson DG, Tomkins JP, Frisch DA, Wing RA, and Paterson AH (2000) Construction of plant bacterial artificial chromosome (BAC) libraries: An illustrated guide. *J. Agric. Genomics* **5**: www.ncgr.org.

*Due to an oversight, USDA postdoctoral award 99-35300-7819 was not acknowledged in these manuscripts. However, Dr. Peterson's contribution to the research was funded by his USDA postdoctoral award.

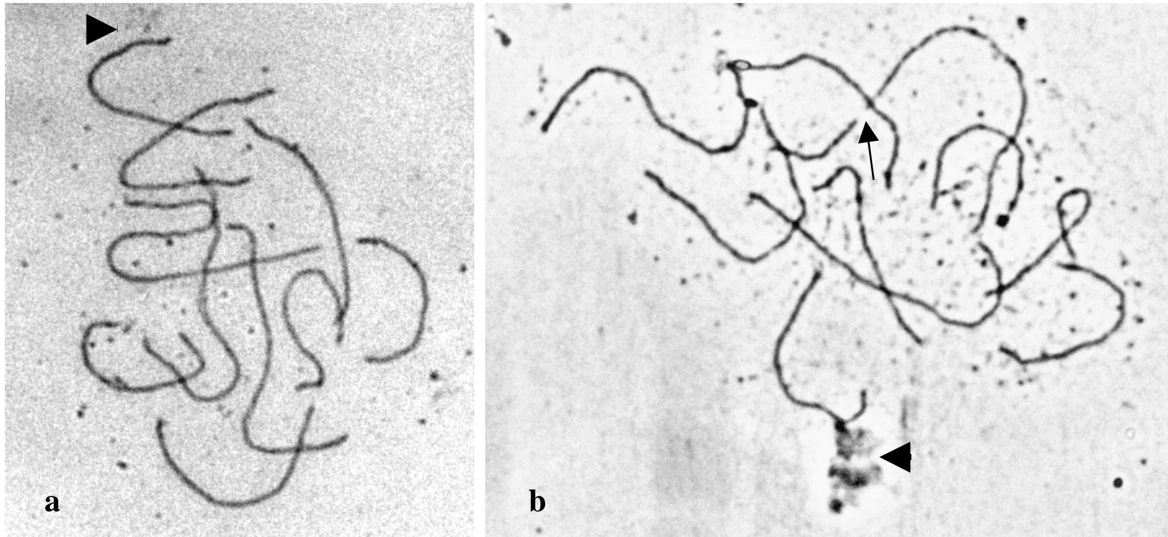

FIGURE 1 – Complete sorghum synaptonemal complex (SC) sets. The SCs were stained with silver nitrate and photographed using (a) bright-field microscopy or (b) phase-contrast microscopy. Magnification = 2500X. In (a) and (b), the sixth longest SC is associated with an amorphous structure of unknown origin (arrowheads). In (b) one of the ten SCs is broken into two parts (an arrow marks the breakpoint).

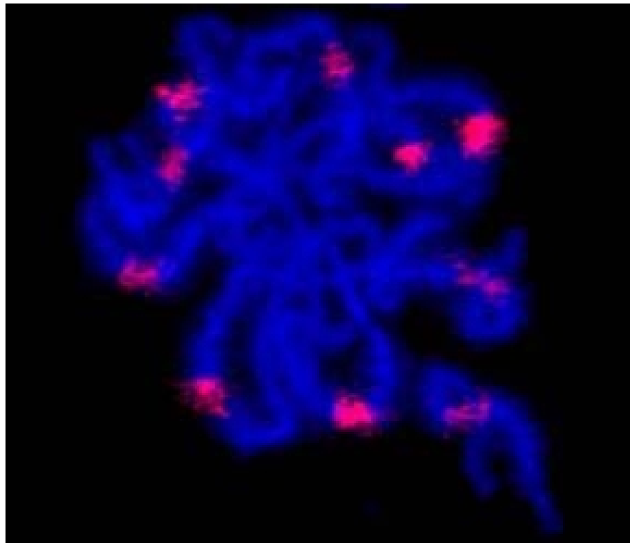

FIGURE 2 – Detection of sorghum centromeres by fluorescence *in situ* hybridization (FISH). The sorghum pachytene chromosomes (SCs) have been counterstained with DAPI (blue fluorescence) while the sites of probe hybridization are marked by rhodamine (red) fluorescence. The probe used in FISH was pSau3A10, a sequence unique to the centromeres of species in the *Sorghum* genus.

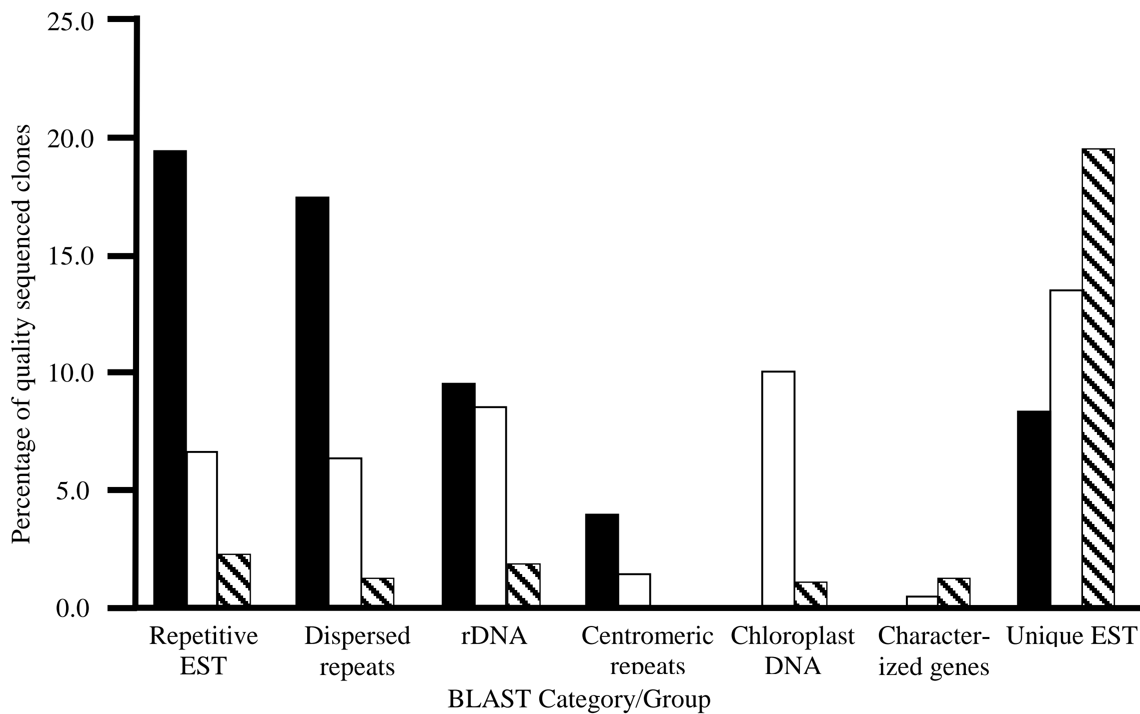

FIGURE 3 – Sequence composition of the different sorghum Cot libraries (based on sequence analysis of 250-500 clones for each library). Black bars represent HRCot clones, white bars represent MRCot clones, and diagonally striped bars represent SLCot clones (see **Peterson et al. 2002b** for details).