

**SURVEY OF 42,000 GOSSYPIUM HIRSUTUM CV. MAXXA BAC-END SEQUENCES
AND FREQUENCY, TYPE, AND ANNOTATION OF BAC-DERIVED SSRs**

Michael B. Palmer

Clemson University

Clemson, SC

James Frelichowski

USDA-ARS

Shafter, CA

Dorrie Main

Clemson University

Clemson, SC

Mauricio Ulloa

USDA-ARS, WICS Res. Unit, Cotton Enhancement Program

Shafter, CA

Roy Cantrell

Cotton Incorporated

Cary, NC

Stephen Ficklin

Jeffrey P. Tomkins

Clemson University

Clemson, SC

Abstract

The quest for more molecular markers is a major initiative in cotton, which lags behind crops such as soybean, maize, and rice in this type of research. In an effort to increase the number of microsatellite markers in *Gossypium*, BAC-end sequences from a publicly available *Gossypium hirsutum* cv. Maxxa BAC library (Tomkins 2001) were mined for microsatellites, or SSRs. Mononucleotide repeats were not included in the analysis. The minimum number of repeats accepted for each motif were as follows: 5 for dinucleotide repeats, 4 for trinucleotide repeats, 3 for tetranucleotide repeats, 3 for pentanucleotide repeats, and 3 for hexanucleotide repeats. BAC clones were sequenced from both ends, tested for redundancy, and screened against the GenBank non-redundant protein and MIPS Arabidopsis databases. Further annotation was performed using the gene ontology terms associated with matching sequences in the SwissProt database, and by performing a scan of the Integrated resource of Protein Families, Domains and Sites (InterPro). The sequences were then submitted to GenBank. The GenBank-submitted sequences were re-analyzed at a higher Phred stringency, and the resultant high-quality sequences were mined for microsatellites.

From 38,000 high quality sequences, approximately 7,000 microsatellites were developed. These sequences were analyzed for the type of repeat motif, GC content, frequency, motif frequency within the total number of microsatellites, and the presence of open reading frames. Primer3 was used to derive primers for the flanking sequences of the SSRs. Dinucleotide and tetranucleotide repeats made up 68% of the microsatellites. The genomic SSRs should improve marker saturation of the cotton genome and allow SSRs to be anchored to genomic clones, aiding in the reconciliation of extant genetic maps to physical maps. The data are stored and accessible from two websites; the Clemson University Genomics Institute (<http://www.genome.clemson.edu/projects/cotton/>) and the Cotton Microsatellite Database (<http://www.cottonssr.org>).