

CHROMOSOMAL LOCATIONS OF KNOWN-FUNCTION ESTs IN COTTON

Peng Chee and Kippy Lewis
Coastal Plain Experimental Station
Tifton, GA
Junkang Rong and Andrew Paterson
University of Georgia
Athens, GA

Abstract

Due largely to the efforts of two cotton EST projects, approximately 21,000 ESTs from *G. arboreum* and another 9,500 ESTs from *G. hirsutum* have been deposited into the Genbank. These ESTs, generated by “single pass” sequencing of randomly selected complementary DNA (cDNA) clones synthesized from expressing gene sequences, are mostly anonymous in that their functions and map locations have not been characterized. With a number of model organisms in both animals (human and mouse) and plants (*Arabidopsis*) whose genome has been completely sequenced, together with the growing list of Express Sequence Tag (EST) database being developed, DNA sequences of known-function and candidate genes have increased exponentially in recent years. The volume of characterized-genes currently available in the Genbank has now permitted researchers to elucidate the function of an anonymous gene sequence by simply cross-referencing it with other known gene sequences in the Genbank through a “blast” database search. The objectives of this project are to conduct a blast search through the Genbank database to identify cotton EST sequences that show homology to well-characterized genes from other species and further, to map these known-function ESTs onto an anchor RFLP map.

A preliminary blast search was carried out using 465 (or roughly 2% of the) *G. arboreum* EST sequences to determine the usefulness of this database for gene discovery and genetic markers development. The blast results revealed that 20% (or 94) of the ESTs contain well-characterized genes, 25% show homology to DNA sequences that have not been characterized (ie. BACs, chromosomal sequences, genomic clones, etc.), and 45% do not show homology to any sequences in the Genbank. Only 10% of the ESTs are of non-useful sequences, which include genes with high copy number such as ribosomal and histone genes, or from the organelle genomes such as the Mitochondria and Chloroplast genomes.

We have developed PCR primers from *G. arboreum* EST sequences (named Express Sequence Tag Sites – or ESTS) that show high homology to well-characterized genes. Of the 94 ESTS that we synthesized, 42 amplified to a single DNA fragment in *G. hirsutum* var. Palmeri and *G. barbadense* (accession K101). Thus far, the PCR products of five out of the 20 ESTS that have been subjected to restriction digest with four enzymes show polymorphisms between K101 and Palmeri. The observed polymorphism rate of 25% (5/20) is lower than other DNA marker systems such as SSR, AFLP and RAPD. This low polymorphism rate is not unexpected as gene sequences are inherently more conserved than the non-genic and often repetitive in nature DNA sequences analyzed by the above mentioned marker systems. The low polymorphism rate is compensated by the higher return in genetic information these ESTS markers provide in that the function of these genes are well understood, and four of the loci are mapped with high confidence (LOD > 4) on an anchor RFLP map. We anticipate that these ESTS markers will be the primary tool in cotton for the transition from genetic linkage mapping to a candidate gene mapping approach for important phenotypes that is now adopted by many organisms.

Acknowledgements: We thank the University of Georgia Research Foundation for funding this research project.