A NEW CLASS OF DNA MARKERS IN COTTON

Pawan Kumar and Peng W Chee
Dept. of Crop and Soil Sciences, UGA, Tifton

Introduction

Molecular markers have wide spread applications in plant improvement. As the name implies, they are "land marks" along a linear stretch of DNA which make up the genome or the total genetic composition of an individual. The application of molecular markers to plant breeding can be divided into three main categories: (1) the characterization of germplasm, known as fingerprinting; (2) the genetic dissection of the target trait, (identification and characterization of genomic regions involved in the expression of the target trait); and, (3) following the identification of the genomic regions of interest, crop improvement through marker-assisted selection (MAS). The first two applications have proven their value by generating knowledge about the genetic diversity of germplasm, thereby allowing placement into heterotic groups and a better understanding of the genetic basis of agronomic traits of interest.

The most common types of molecular markers are based on analyzing anonymous DNA sequences (e.g., RFLP, RAPD, AFLP and microsatellite markers), meaning that the sequences fall on parts of the genome that do not correspond to any functional genes. However, some molecular markers (including cDNA and EST markers, as well as the protein markers) do correspond to functional genes. The limitation of using functional genes as markers is that the level of polymorphisms is generally lower than anonymous markers – most genes do not tolerate any changes in DNA sequence due to mutation. One possible solution to increase the level of polymorphism is to target genic regions that do not translate into protein sequences. These non-transcribed sequences, also known as introns, are common among plant and animal genomes. In this study, our objective is to investigate the possibility of predicting the presence of introns in cotton genes based on analyzing genes that perform the same function in other well studied species such as rice and Arabidopsis.

Materials And Methods

Because the genome of upland cotton is considerably more complex due to polyploidy, we therefore focus our study by analyzing gene sequences from the diploid progenitors, *Gossypium arboreum*.

There are approximately 21,000 ESTs from *G. arboreum* in the Genbank. These ESTs are generated by “single pass” sequencing of randomly selected complementary DNA (cDNA) clones synthesized from expressing gene. Our approach was to first elucidate the function of an cotton gene sequence by simply cross-referencing it with other known gene sequences in the Genbank through a “blast” database search.

We then identify the untranslated regions (UTRs) or the introns of a gene by comparing the cotton gene sequences against the genomic sequences of *Arabidopsis* and/or *Oryza.* Based on the alignment pattern with blastn, only those ESTs which showed "gaps" in the alignment were selected. These gaps were assumed to be putative introns. Finally, the BioEdit v5.0.6 software was used for aligning different sequences. Polymorphism in length and base composition of Introns and the flanking exons was studied.

Results And Discussion

A preliminary blast search was carried out using 465 *G. arboreum* EST sequences to determine the usefulness of this database for gene discovery and genetic marker development. Nineteen EST specific primers were generated and were used to amplify genomic DNA from *G. herbaceum* (A1), *G. arboreum* (A2), *G. raimondii* (D5), *G. trilobum* (D8), *Gossypium kirkii,* and *G. sturtianum* (C1). Subsequently the amplified fragments were both forward and reverse sequenced and the sequences were compared with the EST sequence from TIGR cotton gene index. Out of these 19 primer sets, three primer sets did not amplify the targeted region as no homology exsisted between the sequence of the amplified fragment and the EST.

Out of the remaining sixteen primer sets, thirteen showed the presence of introns. Size of cotton introns varied from 71 base pairs (bp) to 608 bp, when compared to the introns of *Arabidopsis* and/or *Oryza*, in general it is found that cotton has smaller introns.To study the level of diversity in introns across cotton genome, introns from different cotton genomes were compared to *G. arboreum* (A2) and it was found that *Gossypium kirkii* was most divergent and was 85.42% similar to *G. arboreum* (A2). *G. herbaceum* (A1) was genetically most similar with 96.88% similarity to *G. arboreum* (A2). Introns from *Arabidopsis* and *Oryza* showed 34.45% & 25.69% similarity, respectively. Thus, intronic regions were more variable than the flanking exons.

Upon comparison of the alignment of A1 & A2 genomes, a number of single point mutations were evident between the two genomes. Our future goal in this study is to develop a collection of molecular markers targeting these single point mutations for use in genetic mapping experiments.

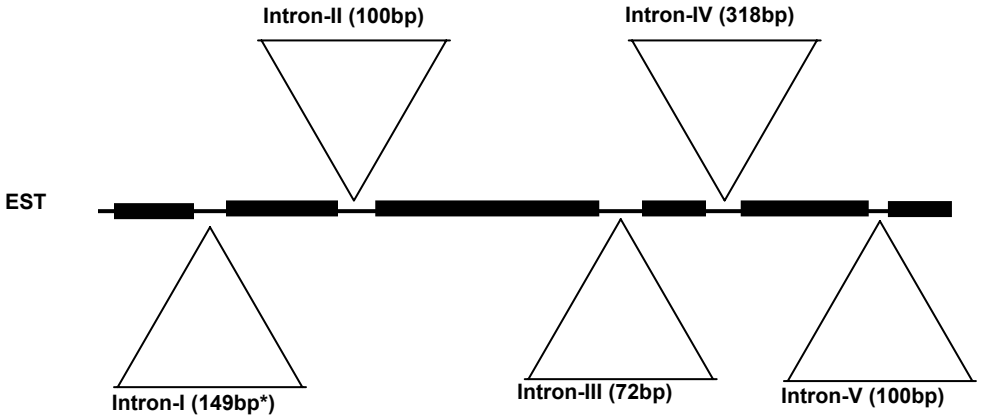

Fig.1 Primer-280 for EST(584bp) putative NADH-ubiquinone oxireductase
Intron-I *Arabidopsis(149), Oryza(121bp)
Intron-II A1(100bp), A2(100bp), D5(101bp), D8(100bp), C1(100bp), Arabidopsis(93bp), Oryza(121bp)
Inrton-III A1(73bp), A2(72bp), D5(73bp), D8(72bp), C1(71bp), Arabidopsis(75bp), Oryza(96bp)
Intron-IV A1(318bp), A2(318bp), D5(316bp), D8(309bp), C1(316), Arabidopsis(105bp), Oryza(86bp)
Intron-V A1(98bp), A2(100bp), D5(100bp), D8(100bp), C1(100), Arabidopsis(88bp), Oryza(101bp)

¥ Data for *Gossypium kirkii* was not available for this primer