

Analysis of Expressed Sequence Tags (ESTs) and Microsatellite Sequences in Coffee (*Coffea arabica*)

M.A. CRISTANCHO¹, A.L. GAITÁN¹, S.R. McCOUCH²

¹National Center of Coffee Research (CENICAFE), Chinchiná, Caldas, Colombia.

²Department of Plant Breeding, Cornell University, Ithaca, NY 14853.

SUMMARY

Despite its global importance, very little information has been gathered from the coffee plant at the genetic level. Here we describe the analysis of a set of Expressed Sequence Tags (ESTs) and microsatellite sequences developed from coffee (*Coffea arabica*).

A non-directional cDNA library was constructed from young coffee leaves tissue. Seven hundred and sixty clones were selected, subjected to single-pass sequencing and identified by sequence similarity searches against sequences in the EST and non-redundant GeneBank databases. Of the 751 sequences, 42% exhibit similarity to known plant genes, 19% to sequences with unidentified function, and 39% represent new gene sequences. Analysis of the identified clones indicated sequence similarity to a broad diversity of genes encoding proteins such as enzymes, structural proteins and regulatory factors.

Additionally, simple sequence repeat (microsatellite) loci were identified in coffee in clones isolated from genomic libraries. Coffee turned out to be among plant species containing a low frequency of microsatellites. Sequence analysis showed that (AC)_n microsatellites in coffee are frequently associated with other microsatellites, mainly (TC)_n and (AT)_n motifs, while (AG)_n microsatellites are not normally associated with other microsatellites. The results also suggest that the coffee genome harbors a low number of poly (AAT) microsatellites, a very common microsatellite motif in other plant genomes. Dot-blot and sequence analysis revealed that a high number of coffee microsatellites are associated to miniature inverted-repeat transposable elements or MITEs.

We have developed nearly 400 microsatellite markers and currently are developing a *C. arabica* molecular map with them. These markers have been also used to examine the genetic diversity of CENICAFE germplasm collection.