

Sequence Analysis from Leaves, Flowers and Fruits of *Coffea arabica* var. Caturra*

G. MONTOYA¹, H. VUONG², C. R. BUELL², M. CRISTANCHO¹, P. MONCADA¹,
M. YEPES³

¹National Coffee Research Center – CENICAFE – Chinchiná, Caldas, Colombia

²The Institute for Genomic Research TIGR

³Cornell University

*Research co-sponsored by the Colombian Ministry of Agriculture and The National Science Foundation Plant Genome Research Program

SUMMARY

Coffee is one of the main agricultural products in Colombia and increased agronomic and genetic knowledge is necessary to establish effective strategies for its improvement. Currently, molecular studies are in progress to identify genes of agronomic importance. In this study, we report on cDNA libraries from three different tissues (leaf, flower, and fruit) from *C. arabica* var. Caturra which were developed to provide data on genes expressed in leaf, flower, and fruit tissues. In total, 32,961 sequences were generated and following cleaning, clustering and assembly, collapsed into 10,799 unique sequences. Putative homologs for the *C. arabica* sequences could be found in Arabidopsis and rice, finished dicot and monocot genomes. Furthermore, putative homologues could be identified in publicly available EST collections of *Coffea canephora* as well as in the Solanaceae species, *Lycopersicon esculentum* and *Solanum tuberosum*. Annotation of the sequences revealed a diversity of genes including similarity to proteins involved in metabolism, protein synthesis, and cellular defense. This distribution agrees with previous studies made in tomato, where transcripts were mainly involved in metabolic processes of transcription and protein synthesis. These categories were also found in Arabidopsis and can represent a conserved set of metabolic functions in plants.

Key words: *Coffea arabica*, cDNA libraries, ESTs, sequence similarity, putative function., genomics