

A Catalogue of Genes Expressed in *Coffea arabica* L.

A. PALLAVICINI, L. DEL TERRA, B. DE NARDI, P. ROVELLI, G. GRAZIOSI

Dipartimento di Biologia, Università di Trieste, Trieste, Italy.

SUMMARY

Plant synthesise more than 100,000 different compounds, many of which have high nutritional value or industrial applications. In the past, the isolation and characterisation of a relevant gene was a complex and time consuming task. More recently the EST (Expressed Sequence Tag) analysis have provided a cost-effective and rapid alternative route to the isolation of large numbers of genes. Knowledge of expressed genes in *Coffea arabica* is practically non existent and therefore we undertook an EST project.

About 30 seeds of bourbon yellow were allowed to germinate on wet blotting paper. The root (about 15 mm) were cut and immediately frozen in liquid nitrogen, Following the grinding of the tissue, the polyadenylated mRNA was extracted and purified by oligo-dT column chromatography by standard techniques. The first strand of cDNA was synthesised by reverse transcriptase and an anchored oligo-dT-NotI primer. After completion of the second strand, using the Gubler and Hoffman strategy, the cDNA fragments were repaired with T4 DNA Polymerase, ligated to BstXI non-palindromic adaptors, digested with NotI and size-fractionated on CLB4 spun column. The cDNA was directionally cloned into a *BstXI*, *NotI* digested pcDNAII plasmid vector (Invitrogen) and the ligation reaction was used for transforming electro-competent *E. coli* DH10B cells. PCR tests were run on a sample of the library to determine its quality. We obtained 800,000 recombinant bacterial clones with an average insert size of 1 kb. More than 500 random clones were picked-up in the 384 wells plates and the inserts amplified by PCR. The positive PCR products were re-arrayed in new plates and sequenced on a 3,700 ABI sequencer.

The first catalogue of gene expressed in the meristematic radical tissue of *Coffea arabica*, shows a large number of genes with a high degree of homology with sequences of other plant species like *A. thaliana*, *L. esculentum*, *O. sativa* etc. About 35% of the sequences showed no homology to known sequences. As expected, we obtained a relatively large number of sequences coding for ribosomal proteins. The second most abundant class of sequences was those coding for chitinases. Most probably these genes are involved in the defense response of the young roots to bacteria and moulds. Other genes with high expression are the extensin and the acid phosphatases.

Besides the abundant classes of expressed genes mentioned above, we found some unique sequences, which are amenable for further developments. This group includes the glycoprotein EP1, a member of S locus protein family, which could be used to study the autofertility of *C. a.*; one of the many resistance genes, presumably a rust resistance gene.