

Generation and Analysis of a Coffee EST Database: Deductions about Gene Repertoire, Expression and Evolution

C. LIN¹, L. A. MUELLER¹, J. McCARTHY², V. PÉTIARD², D. CROUZILLAT²,
S. TANKSLEY¹

¹Departments of Plant Breeding and Plant Biology, Cornell University,
Ithaca, NY 14853, USA

²Nestlé Research Center, Tours, 101, Avenue Gustave Eiffel, B.P.49716,
37097 Tours Cedex 2, France

SUMMARY

An EST database has been generated for coffee, *Coffea canephora* var Robusta, based on sequences from approximately 47,000 cDNA clones derived from 5 different tissues/stages, with a special focus on developing seeds. When computationally assembled, these sequences corresponded to 13,175 unigenes, which are currently being analyzed for functional annotation, differential expression, single nucleotide polymorphisms (SNP), simple sequence repeats (SSR), and complete representation of biochemical pathways of interest. Here, we present a brief overview of how the sequences were generated, assembled, annotated, and compared to Arabidopsis and tomato. A relational database was created to hold the complete datasets, and a web interface was developed for easy access of the database. This infrastructure is a starting point for genomics research in coffee.

RÉSUMÉ

Une base de données d'EST de Caféier, *Coffea canephora* var Robusta, a été créée à partir d'approximativement 47 000 clones d'ADNc de 5 types de tissus ou stade de développement différent plus particulièrement à partir de graines. Après assemblage informatique ces séquences représentent 13 175 unigènes qui sont en cours d'analyse pour leur annotation fonctionnelle, leur expression différentielle, le polymorphisme pour un nucléotide, les microsatellites, et la représentation des voies métaboliques et biochimiques d'intérêt. Nous résumons rapidement ci-après l'obtention des séquences, leur assemblage, leur annotation et leur comparaison avec Arabidopsis et la tomate. Une banque de données relationnelle a été créée pour stocker ces données et un accès internet aisé a été développé. Cette infrastructure constitue un point de départ pour la recherche génomique sur le caféier.