Development of a Bioinformatics Platform at the Colombia National Coffee Research Center

M. CRISTANCHO[1], L. RIVERA[1], C. OROZCO[1], A. CHALARCA[1], L. MUELLER[2]

[1]National Coffee Research Center, CENICAFE, Planalto, Chinchiná, Caldas, Colombia
[2]Plant Breeding Department, Cornell University, Ithaca, NY, 14850

SUMMARY

We have implemented a web-based Bioinformatics platform that functions as a genomics information resource for coffee and other organisms studied at the Colombia National Coffee Research Center – CENICAFE. The Bioinformatics platform includes a Laboratory Integrated Management System (LIMS), the implementation of wEMBOSS, home-developed perl tools for data analysis, InterproScan for annotation of sequence domains, and the implementation of wBLAST and wNetBLAST among other tools available. The main backbone of the system is an adaptation of the SOL Genomics Network (SGN) databases developed at Cornell University for ESTs, molecular markers and BAC sequences storage and analysis (http://sgn.cornell.edu). The system is based on the postgresQL relational database, the use of perl scripts for the manipulation of data, the Apache Web server with the mod–perl integrated perl interpreter, and the servers run the Debian distribution of the GNU/Linux operating system. Although SGN has mainly developed as a plant genomics oriented resource, the Cenicafe platform has implemented several new tools and databases for the analysis of other organisms sequence data such as fungi and insects. The Cenicafe databases contain to date 32.000 coffee EST sequences from 22 libraries organized in 9.257 *C. arabica* and 1.239 *C. liberica* unigenes, 6.000 *Beauveria bassiana* EST sequences organized in 2.404 unigenes, and 4.000 *Hypothenemus hampei* (coffee berry borer) EST sequences organized in 885 unigenes, besides the more than 100.000 Solanaceae unigene sequences annotated at SGN. The sequences are annotated based on Solanaceae, Arabidopsis, Swissprot and Genbank sequence comparisons using BLAST homology searches, aminoacids are predicted using ESTScan, the domains are annotated using InterproScan and Gene families are annotated using a perl script developed at SGN. The system will implement in the near future a database of coffee genetics resources developed at Cenicafe, a proteomics platform, and a Microarray database. We will also be incorporating other components to the platform specially for the visualization of genetic maps from the Gmod project (Gbrowse), the SGN system, TIGR, and other open source projects.