

Poster B-65

CocoaGen DB, an Integrative Information System to Exploit both Phenotypic and Genomic Data on Cocoa.



Authors:

Xavier Argout (*CIRAD*)

Manuel Ruiz (*CIRAD*)

Mathieu Rouard (*INIBAP*)

Chris Turnbull (*School of Plant Sciences, University of Reading*)

Claire Lanaud (*CIRAD*)

Eric Rosenquist (*ARS, USDA*)

Brigitte Courtois (*CIRAD*)

Short Abstract: CocoaGen DB is an integrative information system on cocoa initiated through a collaborative project involving CIRAD, the University of Reading and USDA. This database comprises molecular, genetic, phenotypic and passport data on cocoa. An intuitive web consultation interface with several tools allows end-users to best exploit genetic information available on cocoa germplasm.

Long Abstract:

CocoaGen DB is an integrative information system that comprises molecular genetic, genomic, phenotypic and passport data on cocoa. It was initiated through a collaborative project involving CIRAD (Centre de coopération internationale en recherche agronomique pour le développement, France), University of Reading (School of Plants Sciences, UK) and USDA (United States Department of Agriculture, USA). This database combines molecular genetic information contained in TropgeneDB (<http://tropgenedb.cirad.fr>) with phenotypic and passport data contained in ICGD (International Cocoa Germplasm Database, <http://www.icgd.reading.ac.uk/>).

CocoaGenDB development is mainly based on the AceDB database management system (J. Thierry Mieg and R. Durbin, 1996, <http://www.acedb.org>), running on Linux Redhat 7.1. AceDB is an object-oriented system capable of storing and retrieving complex biological information, and is currently used by many databases such as WormBase. AceDB provides an intuitive object-oriented view of biological data, and a graphical user interface with many specialized data visualization tools, such as a genetic map viewer and a sequence annotation display. Database exchanges between ICGD, TropgeneDB and CocoaGenDB are made using the XML data format. The CocoaGenDB web consultation interface is implemented with Perl/CGI scripts, using modules of the AcePerl API (Application Programming Interface) and the AceBrowser generic web interface.

Most of the ICGD and TropgeneDB data are being integrated into CocoaGenDB. These data include detailed phenotypic information on geographic origins, agro-morphological data, disease, pest and abiotic stress reactions, parentage, over 28 000 accessions. Associated molecular and genetic data comprise the genotypes at various markers (RFLP, AFLP, microsatellites, isozymes, etc.), information on the markers themselves, genetic maps, QTL, genes, etc. AceDB allows us easily including images in CocoaGenDB such as gel profiles,

pictures of fruits or disease symptoms, etc.

CocoaGen DB is available by internet (<http://cocoagendb.cirad.fr/>) via a user-friendly and intuitive web consultation interface that allows performing complex queries combining genetic and phenotypic information; an example of query could be: 'What are the clones resistant to a disease which have specific alleles at a range of loci?'. To display maps and study synteny relationships, the Comparative Map Viewer CMAP was linked to CocoaGen DB. Furthermore, we developed a Java applet tool for the visualization of clones genealogy and for alleles tracking. We created submission forms to simplify the integration of genomic data and to control their quality by the verification of identifiers and datatypes. A tool to run multivariate analyses using the multilocus profile of the accessions at a set of reference microsatellites has been developed with R and will be connected to the database.

The cocoa community is finishing the development, sequencing and treatment of 200 000 cDNA enriched in full length cDNA. The database has been modelled to integrate sequences, clustering, Blast and GO annotation and libraries description of the data coming from this program. Microsatellites and SNPs derived from the analysis of this cDNA collection will also be added to the database.

To sum up, CocoaGen DB was specifically designed to allow end-users (breeders or molecular geneticists) to best exploit genetic information available on cocoa germplasm.