



Plant & Animal Genomes XVI Conference

January 12-16, 2008
Town & Country Convention Center
San Diego, CA

P26 : Genome Sequencing & ESTs

A First Step In The Understanding Of The Cocoa Transcriptome: Analysis Of An Exhaustive Dataset Of ESTs On Theobroma cacao Generated From Various Tissues And Under Various Conditions.

Xavier Argout¹ , Olivier Fouet¹ , Patrick Wincker³ , Karina Gramacho² , Thierry Legavre¹ , Xavier Sabau¹ , Ange Marie Risterucci¹ , Corinne Da Silva³ , Gaston Loor⁴ , Julio Cascardo⁵ , Brigitte Courtois¹ , David Kuhn⁶ , Ray Schnell⁶ , Regis Babin^{7,8} , Olivier Sounigo^{7,8} , Michel Ducamp⁹ , Joseph Verica¹⁰ , Marc Guiltinan¹⁰ , Laurence Alemanno¹¹ , Regina Machado¹² , Wilberth Phillips¹³ , Siela Maximova¹⁰ , Claire Lanaud¹

¹ CIRAD - URM DAP TA40/03-34398 Montpellier cedex 5 – France

² CEPLAC Km 22 Rod. Ilheus Itabuna, Cx. postal 07, Itabuna 45600-00- Bahia BRAZIL

³ GENOSCOPE - 2 rue Gaston Crémieux □CP 5721 □91057 □Evry Cedex - France

⁴ INIAP: EET-Pichilingue. Código Postal 24 Km 5 vía Quevedo El Empalme, Ecuador.

⁵ UESC-Laboratório de Genômica e Expressão Gênica Rodovia Ilhéus-Itabuna, Km 16- Ilhéus-BA- (BRAZIL)

⁶ USDA-ARS, 13601 Old Cutler Rd. Miami, Florida – USA

⁷ IRAD - Nkolbisson, BP 2067 Yaoundé Cameroon

⁸ CIRAD-UPR Bioagresseurs de pérennes TA 80 / 02 - 34398 Montpellier Cedex 5 - France

⁹ CIRAD- UMR BGPI TA41/K 34398 Montpellier France

¹⁰ Penn State University- The Biotechnology Institute-306 Wartik Lab- University Park, PA 16802-5807 USA

¹¹ CIRAD- UMR BEPC TA 80 / 03 - 34398 Montpellier Cedex 5- France

¹² MASTERFOODS Almirante- Brazil

¹³ CATIE P.O.Box 7170, Turrialba, Costa Rica

To better understand the cocoa transcriptome and to provide a framework for future investigations on cocoa functional genomics, approximately 175 000 cocoa clones from 52 cDNA libraries were sequenced. These libraries were constructed with RNA isolated from 13 genotypes (mainly from the two genotypes Scavina6 and ISC1) and from various tissues and organs (flowers, seeds, cherels, pod, shoot, root or embryos) under various conditions (fermentation, abiotic and biotic stresses). Suppression subtractive hybridisation (SSH) libraries were also produced under various plant pathogen interactions and from flowers pollinated by self compatible or incompatible pollen.

From these cDNA libraries, 149 650 good quality ESTs were generated and analyzed using our “in house” bioinformatics pipeline. The assembly process produced 13837 contigs and 66746 singletons that represent 80583 different potential transcripts. These unigenes were assigned with BLASTX and BLASTN similarities, Gene Ontology annotation and protein domains. A comparison with public ESTs cocoa sequences already available in public domain indicates that 80% of the unigenes may not have been described in public databases.

This assembly and information associated will be available throw the international Web portal CocoaGenDB (<http://cocoagendb.cirad.fr>). A detail analysis of this unigene dataset will be presented.