Plant & Animal Genomes XIX Conference

January 15-19, 2011
Town & Country Convention Center
San Diego, CA

W107: Cacao Genome Sequencing & Cacao Breeders

Deciphering The Genome Structure Of Theobroma cacao

Xavier Argout[1] , Jerome Salse[2] , Jean-Marc Aury[3,4,5] , Mark Guiltinan[6] , Gaetan Droc[1] , Jerome Gouzy[7] , Mathilde Allegre[1] , Cristian Chaparro[8] , Thierry Legavre[1] , Siela Maximova[6] , Michael Abrouk[2] , Florent Murat[2] , Olivier Fouet[1] , Julie Poulain[3,4,5] , Manuel Ruiz[1] , Yolande Roguet[1] , Maguy Rodier-Goud[1] , Jose F. Barbosa-Neto[8] , Francois Sabot[8] , Dave Kudrna[9] , Jetty Ammiraju[9] , Stephan Schuster[10] , John Carlson[11,12] , Erika Sallet[7] , Thomas Schiex[13] , Anne Dievart[1] , Melissa Kramer[14] , Laura Gelley[14] , Zi Shi[15] , Aurelie Berard[16] , Christopher Viot[1] , Michel Boccara[1] , Ange-Marie Risterucci[1] , Valentin Guignon[1] , Xavier Sabau[1] , Michael Axtell[17] , Zhaorong Ma[17] , Yufan Zhang[14] , Spencer Brown[18] , Mickael Bourge[18] , Wolfgang Golser[9] , Xiang Song[9] , Didier Clement[1] , Ronan Rivallan[1] , Mathias Tahi[19] , Joseph Moroh Akaza[19] , Bertrand Pittolat[1] , Karina Gramacho[20] , Angelique D'Hont[1] , Dominique Brunel[16] , Diogenes Infante[21] , Ismael Kebe[18] , Pierre Costet[22] , Rod Wing[9] , William McCombie[14] , Emmanuel Guiderdoni[1] , Francis Quetier[23] , Olivier Panaud[8] , Patrick Wincker[3,4,5] , Stephanie Bocs[1] , Claire Lanaud[1]

[1] CIRAD - Biological Systems Department – UMR DAP TA A 96/03- 34398, Montpellier, cedex 5- France
[2] Institut National de la Recherché Agronomique UMR 1095, 63100 Clermont-Ferrand, France
[3] CEA, IG, Genoscope, 2 rue Gaston Crémieux CP5702, F-91057 Evry, France
[4] CNRS, UMR 8030, 2 rue Gaston Crémieux, CP5706, F-91057 Evry, France
[5] Université d'Evry, F-91057 Evry, France
[6] Penn State University, Department of Horticulture and the Huck Institutes of the Life Sciences, University Park, PA

16802, USA

7 INRA-CNRS LIPM Laboratoire des Interactions Plantes Micro-organismes, BP 52627, 31326 Castanet Tolosan Cedex, France

8 UMR 5096 CNRS-IRD-UPVD, Laboratoire Génome et Développement des Plantes, Université de Perpignan, 52 Avenue Paul Alduy, 66860 Perpignan Cedex, France

9 Arizona Genomics Institute and School of Plant Sciences, University of Arizona, Tucson AZ 85721, USA

10 Penn State University, Department of Biochemistry and Molecular Biology, University Park, PA 16802, USA

11 Penn State University, The School of Forest Resources and the Huck Institutes of the Life Sciences, University Park, PA 16802, USA

12 The Department of Bioenergy Science and Technology (WCU), Chonnam National University, 333 Yongbongro, Buk-Gu, Gwangju, 500-757, Korea

13 Unité de Biométrie et d'Intelligence Artificielle (UBIA), UR875 INRA, F-31320 Castanet Tolosan France

14 Cold Spring Harbor Laboratory, NY 11723, USA

15 Penn State University, Plant Biology Graduate Program and the Huck Institutes of the Life Sciences, University Park, PA 16802, USA

16 INRA, UR 1279 Etude du Polymorphisme des Génomes Végétaux, CEA Institut de Génomique, Centre National de Génotypage, 2, rue Gaston Crémieux, CP5724, 91057 Evry, France

17 Penn State University, Bioinformatics and Genomics Ph.D. Program &Department of Biology, University Park, PA 16802, USA

18 Institut des Sciences du Végétal, UPR 2355, CNRS, 91198 Gif-Sur-Ivette, France

19 Centre National de la Recherche Agronomique (CNRA), B.P. 808, Divo, Côte d'Ivoire

20 Comissão Executiva de Planejamento da Lavoura Cacaueira (CEPLAC), Km 22 Rod. Ilheus Itabuna, Cx. postal 07, Itabuna 45600-00, Bahia, Brazil

21 Centro Nacional de Biotecnología Agrícola, Instituto de Estudios Avanzados (IDEA), Caracas 1015-A, Venezuela

22 Chocolaterie VALRHONA, 8, quai du général de Gaulle, 26600 Tain l'Hermitage, France

23 Département de Biologie, Université d'Evry Val d'Essonne, 25 boulevard François Mitterrand, 91025 Evry, France

Theobroma cacao L., is a diploid tree fruit species, originated from the South American rainforests, and constitutes an important source of incomes for farmers of tropical countries. We produced a high quality draft genome sequence corresponding to a 16,7X genome coverage of a criollo genotype. The assembly corresponds to 76% of the estimated genome size of the T. cacao genotype B97-61/B2 (430 Mbp). This assembly appears to cover a very large proportion of the euchromatin of the T. cacao genome, allowing to recover 97.8% of the unigene resource (38,737 unigenes assembled from 715,457 EST sequences) in the genome assembly. Annotations revealed 28,798 protein-coding genes among which 82% could be anchored in a high density genetic map. Only 20% of the genome consisted in transposable elements, a significantly lower percentage compared to other genome of similar size.

This first cocoa genome sequence was the support for several genome analyses revealing specific extension of some gene families. The comparative mapping of genes involved in disease resistance and quality traits, localised in the genome sequence, and QTLs related to these traits highlighted several co-localisations between them and candidate genes potentially involved in useful cocoa trait variations.

This genome sequence will facilitate a better understanding of trait elaboration and will accelerate T. cacao breeding through efficient marker assisted selection and exploitation of genetic resources.

A genome browser allows to access freely to the cocoa sequence data at the following website : http://cocoagendb.cirad.fr.