



## Plant & Animal Genomes XIX Conference

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

---

P050: Genome Sequencing & ESTs

---

### Sequencing The Cocoa criollo Genome: An International Initiative Of The ICGS (International Cocoa Genome Sequencing) Consortium

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

<sup>1</sup> CIRAD - Biological Systems Department – UMR DAP TA A 96/03- 34398, Montpellier, cedex 5- France

<sup>2</sup> Institut National de la Recherche Agronomique UMR 1095, 63100 Clermont-Ferrand, France

<sup>3</sup> CEA, IG, Genoscope, 2 rue Gaston Crémieux CP5702, F-91057 Evry, France

<sup>4</sup> 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 ( $2n = 2x = 20$ ), originated from the South American rainforests. A criollo genotype, providing a high quality chocolate and highly homozygous, has been choosing by the ICGS to sequence the cocoa genome.

A combination of Sanger and NGS technologies was chosen for sequencing the criollo genotype, with a genome coverage of 16,7 X.

The assembly corresponds to 76% of the estimated genome size of the criollo genotype (430 Mb) and appears to cover a very large proportion of the euchromatin of the *T. cacao* genome, allowing to recover 97.8% of the transcriptome unigene resource. Annotations revealed 28,798 protein-coding genes.

This first cocoa genome sequence was the support to inventory in the cocoa genome important gene families potentially involved in resistance and quality traits. Their localisation in the genome and their comparison with QTLs involved in these traits provide a large set of candidate genes.

The inferred paleohistory of the *T. cacao* genome has highlighted the close evolutionary relationship of the *T. cacao* genome to the eudicot putative ancestor, as was also observed in grape. We propose an evolutionary scenario whereby the ten *T. cacao* chromosomes were shaped from an ancestor through only eleven chromosome fusions. It represents a new and simple model to study evolutionary processes, gene functions, genetics and biochemistry of tree fruit crops.

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