Report International Coffee Genomics Network (ICGN) Meeting and Coffee Genomics Workshop held at the XIX Plant and Animal Genome (PAG) Meeting San Diego, California

January 16-17, 2011

ICGN Workshop Speakers (see abstracts of presentations included at the end of this report):

1. ***Patrick Wincker***, GENOSCOPE-CEA (pwincker@genoscope.cns.fr)
2. ***Luiz Filipe Pereira***, EMBRAPA-Café- IAPAR (Brazil) (filipe.pereira@embrapa.br)
3. ***Carmenza Góngora***, CENICAFE, (Carmenza.Gongora@cafedecolombia.com)
4. ***Valérie Poncet***, IRD (romain.guyot@ird.fr)
5. ***Alvaro Gaitán***, CENICAFE (Alvaro.Gaitan@cafedecolombia.com)
6. ***Girma Agudna***, Ethiopian Institute of Agricultural Research, Jimma Research Center, Ethiopia. (girma.adugna@yahoo.com)

ICGN Meeting invited Speaker

Lukas Mueller, Boyce Thompson Institute (BTI) (lam87@cornell.edu)
Presentation on the *Solanaceae* Genome Network (SGN) www site http://solgenomics.net

Other participants ICGN meeting and/or coffee genomics workshop

1. ***Elijah Gichuru*** Coffee Research Foundation (CRF), Kenya (ekgichuru@yahoo.com)
2. ***Marcela Yepes***, Cornell University (my11@cornell.edu)
3. ***Herb Aldwinckle***, Cornell University
4. ***Aureliano Bombareli,*** Boyce Thompson Institute (ab782@cornell.edu)

5. ***Chifumi Nagai***, Hawaii Agriculture Research Center (HARC) (cnagai@harc-hspa.com)
6. ***Ratnesh Singh***, Hawaii Agriculture Research Center (HARC) (rsingh@harc-hspa.com)
7. ***Ray Ming***, University of Illinois at Urbana-Champaign (rming@life.uiuc.edu)
8. ***David Galbraith,*** University of Arizona, Tucson.
9. ***Jeff Scott***, Purdue University, Illinois
10. ***Nicolas Roux*** Bioversity (n.roux@cgiar.org)
11. ***Romain Guyot***, IRD (Romain.guyot@ird.fr)
12. ***Alexandre De Kochko***, IRD (dekochko@ird.fr)
13. ***James McCarthy***, Nestlé (James.McCarthy@rdto.nestle.com)
14. ***Giovanni Giuliano***, Italian National Agency for New Technologies, Energy and Sustainable Development ENEA (giovanni.giuliano@enea.it)
15. ***Gisella Orjeda***, Universidad Peruana Cayetano Heredia, Perú (gisella.orjeda@upch.edu.pe)
16. ***Adriana Muñoz***, University of Ottawa (amunoz@site.uottawa.ca)
17. ***Seth Findley***, University of Missouri (findleyse@missouri.edu)
18. ***Shannon Straub,*** Oregon State University (straubs@science.oregonstate.edu)
19. ***Javier Terol*** (terol_javalc@gue.es)
20. ***Yoshikazu Tanaka***, Suntory Holdings Limited, Research Center, Chief Executive, Institute for Plant Science, Japan, (Yoshikazu_Tanaka@suntory.co.jp)

Coffee Genomics Workshop at PAG January 16, 2011

Approximately 50 scientists participated in our 4th coffee genomics workshop held as part of the Plant and Animal Genome Meeting in San Diego January 15-20, 2011. The co-organizers of the workshop Philippe Lashermes, Marcela Yepes and Rod Wing thank the speakers for their participation and contributions.

Program and abstracts of workshop presentations and posters on coffee are included as an appendix at the end of this report. We received very positive feedback from several coffee workshop participants to continue the organization of this event during future PAG meetings (http://www.intl-pag.org/). All ICGN members are invited to participate in the 5th Coffee Genomics Workshop that will be held January 15, 2012 as part of the XX PAG meeting in San Diego, January 14-18, 2012. Please contact one of the organizers if interested on presenting a talk or poster, or with suggestions for new topics for workshop presentations or for the round table discussion. The coffee genomics workshop is an excellent opportunity to present advances in coffee genomics research to the International Plant and Animal Genomics Community and is helping our community explore new collaborations as well as funding opportunities.

Report ICGN meeting January 17, 2011

Update ICGN activities, survey, and www page re-design

Following the ASIC meeting in Indonesia, the ICGN Steering Committee held a conference call and the minutes of that teleconference will be posted with this report in our www site. The Steering Committee has contacted Bioversity who has kindly agreed to continue hosting our www site at no-cost. We will be working with them on re-designing the site (http://www.coffeegenome.org) as the coffee genome sequences become available to facilitate exchange of information. Please send suggestions on what type of information you would like to see incorporated in the site to either the ICGN secretariat (Marcela Yepes, my11@cornell.edu) or to the Chair of the Steering Committee (Philippe Lashermes, philippe.lashermes@ird.fr). ICGN will conduct also a survey with all members to establish projects of interest to our community as the coffee reference genomes become available. We look forward to your input to prioritize the projects we will focus on over the next 5 years. ICGN has currently 100 members on its mailing list, representing 52 countries (the majority coffee producing countries). The survey will be conducted at a regional level to enhance global participation of institutes and scientists as well as stakeholders interested on supporting coffee genomics research.

Update Status High-Density Mapping of the diploid species *Coffea canephora*

A high density reference genetic map for *Coffea canephora* Pierre is being constructed in collaboration with R&D Nestlé Center and the Indonesian Coffee and Cocoa Research Institute. The population being mapped is from a cross between two highly heterozygous genotypes, a Congolese group genotype (BP409) and a Congolese-Guinean hybrid parent (Q121). The segregating population is composed of 93 F1 individuals. ICGN members will continue to saturate the map by mapping approximately another 1,000 sequence based markers such as SSRs and SNPs into the Indonesian population. DNAs from the two parental clones (BP409 and Q121) and the segregating progeny (93 individuals) were sent to:

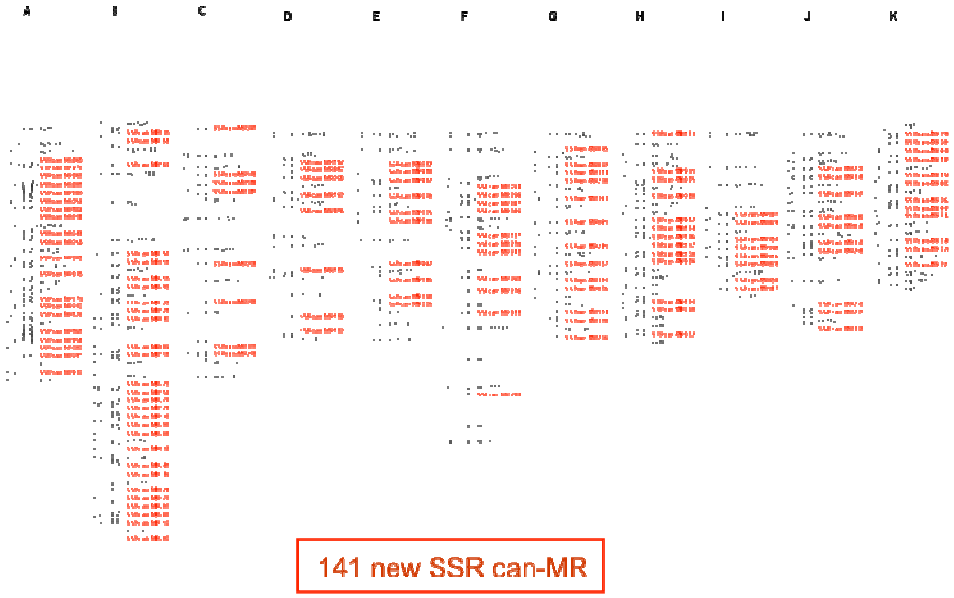
- Ray Ming (University of Illinois, Urbana-Champaign, UIUC, USA)
- Philippe Lashermes (IRD, France)
- Eveline Caixeta (Embrapa, Brazil)
- Gisella Orjeda (Universidad Peruana Cayetano Heredia, UPCH/Perú) who has recently joined the mapping initiative.

R&D Nestlé (Tours, France) provided an anchor genetic map with a set of 115 mapped loci selected to allow the mapping of new SSRs markers from BAC ends.

About 1,500 of theses new SSRs (from IRD BAC-end sequences analysis) were screened for polymorphism within the two parents by IRD, UIUC, and Nestlé. A first set of 141 new SSR makers has been mapped by IRD with the help of Nestlé. R&D Nestlé is currently mapping another set of approximately 150 new SSRs.

The *C. canephora* high-density genetic map will be upgraded according to the progress made by UIUC, Embrapa and UPCH partners. Ideally for genome assembly, this reference coffee genetic map would need to have 2 sequence-based markers per 1 million bp. This map will be used to anchor the scaffolds of the *C. canephora* sequences that are being generated, in order to reconstruct the chromosomal molecules. During the next few months, the mapping efforts will focus on the identification and mapping of SSRs from the *C. canephora* sequence scaffolds that are not or insufficiently anchored. Both, the high-density genetic map as well as the marker information's will be freely available on a dedicated web-site (e.g. SGN web site).

Please send updated data or information and comments to Dominique Crouzillat, Nestlé (dominique.crouzillat@rdto.nestle.com) or Philippe Lashermes (philippe.lashermes@ird.fr). Enclosed is a figure of the International Robusta map as of January 2011 (included in red are the SSRs mapped by IRD).



Update on the status of the *Coffea canephora* genome sequencing

With funding from the French Agency ANR (Agence Nationale de la Recherche), several institutes (Genoscope-CEA, IRD and CIRAD) are combining their

scientific resources and expertise to sequence, assemble, and annotate the entire genome of *C. canephora*. Several other ICGN members have joined these efforts particularly for mapping, genome sequencing and annotation. The *C. canephora* genome consists of 11 chromosomes, is about 710 Mb in size, and is being sequenced *de novo* with deep coverage using different sequencing platforms to obtain a reference genome for *Coffea*. Genoscope is leading the sequencing and assembly. Patrick Wincker, Head of Sequencing and Coordinator of Eukaryote Annotation and Analysis at Genoscope, presented the sequencing strategy and the status of the project during the coffee genomics workshop. Their group was involved in the cacao genome sequencing that was recently completed. The cacao manuscript can give an idea of the results that will be obtained for the sequencing of *C. canephora* (Argout *et al*. 2011. The genome of *Theobroma cacao*. Nature Genetics 43: 101-108; full text at http://www.nature.com/ng/journal/v43/n2/full/ng.736.html)

C. canephora is one of the ancestral progenitors of the widely cultivated, *C. arabica* a recent allotetraploid species formed of the merge of the diploid species *C. canephora* and *C. eugenioides*. The *C. canephora* accession DH200-94, a doubled haploid genotype was selected for sequencing because of its homozygous status to facilitate genome assembly. *De novo* genome sequencing with deep coverage is being performed using both 454 Roche and Illumina technologies. Direct whole genome shotgun sequencing and paired-end sequencing of large insert libraries are underway; two 8kb and 20 kb insert libraries have been constructed. Furthermore, clones from two *C. canephora* BAC libraries were BAC-end sequenced by Genoscope using Sanger technology. At this stage and using only the 454 sequences (single or pair read), the following results have been obtained for *Coffea canephora*:

Size: 540 Mb (76% estimated size)
N50 contigs : 6.2 kb
Coverture 15x
N50 scaffolds : 443 kb
Largest scaffold : 12.8 Mb

More sequencing is therefore required using Roche 454 (20X is planned) and Illumina sequencing deep coverage (50X) to correct 454 sequencing errors.

Update of the *Coffea eugenioides* genome sequencing

This project is funded by FONTAGRO/SECCI, and will be started in 2011. The project will be developed in coordination by CENICAFE and Cornell University. The funding for this project was allocated in December 2010.

Once the reference genomes of the diploid species *Coffea canephora* and *Coffea eugenioides (*parental diploid ancestors of the allotetraploid species *Coffea arabica)* become available, they will serve as frame for sequencing and assembly of the allotetraploid species *C. arabica*, the main cultivated coffee species around the world. ICGN goal by sequencing the coffee genome is to build a solid foundation for deciphering the genetic and molecular bases of important biological traits in coffee that are relevant to growers, processors, and consumers. This knowledge is fundamental to allow efficient use and conservation of coffee genetic resources, and for the development of improved cultivars in terms of quality and reduced economic and environmental costs, as well as, to advance efforts to adapt the crop to climate change.

To ensure a full benefit of the generated coffee genomic sequences and resources by the coffee sector, ICGN will explore through International Funding Agencies, ICO, and the private sector ways of securing funding for long-term maintenance of the coffee genome databases, and for the development of friendly end-user tools as well as to organize training courses to promote community annotation efforts.

Presentation Bioinformatics Web Portal *Solanaceae* Genomics Network (SGN)

Lukas Mueller, Coordinator of the SOL Genomics Network (SGN), a site for genome data for Solanaceae species such as tomato, potato and pepper, kindly presented to our ICGN meeting participants how the bioinformatics is being organized by SGN. The Solanaceae SGN www site (http://solgenomics.net) is as a robust system to store sequences, build gene families to identify orthologs in different species, align phenotypic data and genotypic data, as well as metabolic data. Lukas illustrated how information from molecular genetic maps, physical maps, genotyping, and phenotyping is included in their www site. Since several *Solanaceae* species (including tomato and potato) have been or are being sequenced using next generation sequencing technologies a new browser is being developed for comparative studies within *Solanaceae* (that will expand the current G-Browse to a syntheny browser SYN-Browse). Lukas also described for us the ontology browsers, as well as how the community annotation efforts are being organized through their www site with more than 100 community curators. A genotyping SNP platform is being developed using next generation sequencing technologies as part of the SOL-CAP project (target 8,000 SNPs for potato and 8,000 SNPs for tomato). Several of the resources that SGN has developed are of interest for our community for comparative studies as the coffee genome sequences become available.

Upcoming meetings relevant to our community

SOL meeting in Japan, October 16-20, 2011 (http://www.sol2011.jp)
PAG meeting in San Diego Jan 14-18, 2012 (http://www.intl-pag.org)
ASIC meeting in Costa Rica 2012 (http://asic2012costarica.org)

APPENDIX

- **Pictures from ICGN and workshop meeting**
- **Abstracts from workshop presentations and posters related to coffee**



Appendix

Program and Abstracts Coffee Genomics Workshop 2011

Sunday Afternoon, 16 January 2011 3:50 pm to 6:00 pm

Coffee Workshop - Sunset

Co-Organizers: Philippe Lashermes, L'Institut de recherche pour le développement (IRD), France

(philippe.lashermes@mpl.ird.fr)
and
Marcela Yepes, Cornell University
(my11@nysaes.cornell.edu)
and
Rod Wing, University of Arizona
(rwing@Ag.arizona.edu)

Speakers:

03:55 Patrick Wincker, Genoscope/CEA (France) (pwincker@genoscope.cns.fr) **"Sequencing the Coffee Genome (*Coffea canephora*) "**

04:15 Luiz Filipe Pereira, EMBRAPA-Café/IAPAR (Brazil) (filipe.pereira@embrapa.br) **"Analysis of Nucleotide Diversity in *Coffea* spp. "**

04:35 Carmenza Góngora, Colombian National Coffee Research Center (CENICAFE) (Carmenza.Gongora@cafedecolombia.com) **"Gene Expression of the Coffee Berry Borer (*Hypothenemus hampei*) in Response to Infection by the Entomopathogenic Fungus *Beauveria bassiana*"**

04:55 Valérie Poncet, UMR DEADE, IRD (France) (valerie.poncet@ird.fr) **"Isolation Of Two Novel Ty1-Copia Retrotransposons Nana And Divo From Coffee Trees And Their Usefulness To Reveal Evolutionary Relationships In *Coffea* Genus (*Rubiaceae*)"**

05:15 Alvaro Gaitán, Colombian National Coffee Research Center (CENICAFE) (Alvaro.Gaitan@cafedecolombia.com) **"An Updated Annotation of NBS Domains in *Coffea arabica*"**

05:35 Girma Adugna, Ethiopian Institute of Agricultural Research (EIAR) (girma.adugna@yahoo.com) **"Opportunities and Challenges for *Coffea***

arabica **Production and Germplasm Conservation in its Center of Origin-Ethiopia"**

Sequencing The Coffee (*Coffea canephora*) Genome

Patrick Wincker[1] , **Victor A. Albert**[2] , **Alan A. Andrade**[3] , **Xavier Argout**[4] , **Benoit Bertrand**[4] , **Alexandre de Kochko**[5] , **Giovanni Giuliano**[6] , **Giorgio Graziosi**[7] , **Robert Henry**[8] , **Jayarama Jayarama**[9] , **Philippe Lashermes**[5] , **Ray Ming**[10] , **Chifumi Nagai**[11] , **Steve Rounsley**[12] , **David Sankoff**[13]

[1] Genoscope, Institut de Génomique du CEA, CP5706 91057 Evry cedex France
[2] Department of Biological Sciences, University at Buffalo (SUNY), USA
[3] EMBRAPA, Final Av. W/5 Norte Brasília-DF 70770-917, Brasil
[4] CIRAD 34398 Montpellier Cedex 5, France
[5] IRD Centre de Montpellier BP 64501 34394 Montpellier cedex 5, France
[6] ENEA, Post Box 026, via Anguillarese 301 - 00123 S.M. di Galeria, Roma, Italy
[7] University of Trieste. P.le Valmaura . 34100 Trieste, Italy
[8] University of Queensland. Hartley Teakle Building [#83] St Lucia 4072 Australia
[9] CCRI, Coffee Board, Chikmagalur Dist., Karnataka State, India
[10] University of Illinois at Urbana-Champaign. Urbana, IL 61801 USA
[11] HARC PO Box100, Kunia HI 96759-0100 USA
[12] Dow Agrosciences, Indianapolis, IN, USA
[13] University of Ottawa, 585 King Edward Avenue, K1N 6N5 Ottawa, Canada

Commercial coffee production relies mainly on two closely related species: *Coffea arabica* and *C. canephora*, which account respectively for 70 and 30% of the coffee production. All coffee species are diploid (2n=2x=22) and generally self-incompatible, except for *C. arabica* which is the only tetraploid (2n=4x=44) and self-fertile. Molecular analyses (Lashermes et al. 1999) have indicated that *C. arabica* is a recent allotetraploid (CE genome) formed by hybridisation between two related diploid species: *C. canephora* (C genome) and *C. eugenioides* (E genome). In spite of the close relationship between the two constitutive sub-genomes, *C. arabica* displays diploid-like meiotic behavior with bivalent formation (Krug and Mendes 1940, Lashermes et al. 2000). The genomes of coffee species (Cros et al. 1995; Noirot et al. 2003) appear to be of rather low size (i.e. about 660, 710 and 1300 Mb for *C. eugenioides*, *C. canephora* and *C. arabica*, respectively). Several institutes are combining their scientific resources and expertise to sequence, assemble, and annotate the entire genome of *C. canephora*.

The *C. canephora* genome consists of 11 chromosomes, is about 710 Mb in size, and is being sequenced *de novo* with deep coverage using 454 paired-end and single reads, and 50x coverage with Illumina GAIIx data to obtain a reference genome for *Coffea*. The overall sequencing strategy as well as progress of the project will be described.

Analysis Of Nucleotide Diversity In *Coffea* spp.

Luiz Filipe P Pereira[1,2] , **Karina Yanagui**[2] , **Lucia P Ferreira**[2] , **Douglas S Domingues**[2] , **Luiz G E Vieira**[2] , **David Pot**[1,3]

[1] Embrapa Café,PqEB s/n, Brasília, DF,70770-901, Brasil
[2] Lab. Biotecnologia, IAPAR, Rod. Celso Garcia CID Km 375, Londrina, PR, 86001-970, Brasil
[3] CIRAD, Avenue Agropolis, TA-A-96/03, 34398 Montpellier Cedex, France

Marker-assisted selection becomes a reality for many crops; in perennial crops, the utilization of molecular markers in breeding programs can speed up genotype selection. However, the most important commercial coffee species - allotetraploid *C. arabica* - has a restrict number of available polymorphic markers, which is a consequence of the narrow genetic basis and low molecular variability among cultivars. In order to study the nucleotide diversity in *C. arabica*, as well in other diploid *Coffea* relatives, we sequenced PCR amplified fragments of nine genes in 20 *Coffea* genotypes: twelve *C. arabica*, including eight wild genotypes and four commercial cultivars; and eight *C. canephora* genotypes. Genotypes of *C. eugenioides*, *C. racemosa* and *Psylanthus bengalensis* were also included in this analysis. From a total of 9 Kb analyzed, we found 573 polymorphisms: 500 SNPs; 39 INDELs and 34 SSRs. In *C. canephora* genotypes, we detected 188 polymorphisms (frequency of 2.09/100bp). For *C. arabica* we obtained similar results: 144 polymorphism (frequency of 2.13/100bp). Most of the polymorphism found in *C. arabica* only reflected the differences between ancestral homeologs, and they were monomorphic among different genotypes. However, 19 % of these polymorphisms (27 SNPs) were interespecific for *C. arabica*, and 13 of them were fixed among genotypes. The strategy of this work reflects the importance in using a more diverse panel of genotypes in order to identify SNPs in *C. arabica*, pointing out that the exploitation of wild germplasm will be an important source of genetic variability.

Gene Expression Of The Coffee Berry Borer (*Hypothenemus hampei*) In Response To Infection By The Entomopathogenic Fungus *Beauveria Bassiana*

Carmenza Gongora[1] , Javier Mantilla[1] , Luis-Fernando Rivera[1] , Marco Cristancho[1] , Pablo Benavides[1] , Lucio Navarro[1] , Herb Aldwinckle[2] , Marcela Yepes[2] , David Galbraith[3] , Cheryl Vanier[4] , Alvaro Gaitan[1]

[1] Colombian National Coffee Research Center –CENICAFE, Federación Nacional de Cafeteros de Colombia, FEDERACAFE-Chinchiná, Caldas, Colombia
[2] Department of Plant Pathology and Plant-Microbe Biology, Cornell University, Geneva, New York, USA
[3] Department of Plant Sciences, University of Arizona, Tucson, Arizona, USA
[4] Department of Biological Sciences and School of Life Sciences, University of Nevada, Las Vegas, USA

The coffee berry borer (CBB), *Hypothenemus hampei* (Ferrari) (Coleoptera: Curculionidae: Scolytinae), is the major insect pest of coffee worldwide. The entomopathogenic fungus *Beauveria bassiana* plays a major role in biological control of CBB in Colombia. To understand the interaction between the insect and the fungus, and enhance the efficiency of biological control strategies, we developed a CBB oligoarray containing 5354 probes obtained from cDNA libraries and 454 transcriptome sequencing. The oligoarray plaques were hybridized with 36 independent samples of CBB total RNA which had or had not been infected with the fungus *B. bassiana* 24, 48, or 60 hours prior to sampling, with 6 replicates. Data were normalized within slides, ANOVA was used for differential expression analysis, and permutation-based P-values were adjusted. Out of 5354 probes, 48.6% showed no significant treatment or treatment by time interactions. The response to infection (as compared to non-infected tissue) stayed constant across all time points for 29.9% of probes. The remaining 21.5% had significant differences across time points, with very rare changes in the direction of the infection response within time points (3.2%). Fungal infections suppressed expression for 24.6% of probes, and increased it for 22.6%. Among the pathways affected by the treatment or treatment by time interactions were some related to immune responses and muscle function, cuticle synthesis, hematopoiesis, clathrin-dependent cellular processes, endocytosis, cytochrome oxidases, and antibacterial response. Identification of the metabolic changes under entomopathogen stress provides new target sites for improving current management strategies for the major insect pest of coffee.

Isolation Of Two Novel Ty1-Copia Retrotransposons Nana And Divo From Coffee Trees And Their Usefulness To Reveal Evolutionary Relationships In *Coffea* Genus (Rubiaceae)

Valérie Poncet[1] , Pierre-Olivier Duroy[1] , Christine Tranchant[1] , Paulo Mafra D'Almeida Costa[1] , Caroline Duret[1] , Norosoa Josiane Razafinarivo[1] , Claudine Campa[1] , Emmanuel Couturon[1] , Serge Hamon[1] , Romain Guyot[2] , Alexandre de Kochko[1] , Perla Hamon[1]

[1] UMR DIADE, IRD, Centre IRD de Montpellier, BP 64501, Montpellier Cedex 5, France
[2] Plant Genome and Development Laboratory (UMR 5096 – CNRS – IRD – UPVD), Institut de Recherche pour le Développement (IRD), BP 64501, 34394 Montpellier Cedex 5, France

LTR retrotransposons are mobile genetic elements that contribute significantly to the genetic diversity. Molecular markers based on LTR retrotransposon insertions are commonly used in evolutionary studies to resolve recent and complex phylogenetic relationships. In depth study of the molecular phylogeny of 86 species of *Coffea* genus revealed large unresolved branches, suggesting the necessity to use new variable and informative markers. Two different Ty1-copia LTR retrotransposons, named Nana and Divo, were isolated in *Coffea canephora* using BAC clone sequencing. The two elements that belong to separate Ty1-copia families present similar copy number in the *C. canephora* genome. Sequence-specific amplification polymorphism (SSAP), Retrotransposon amplified microsatellite polymorphism (REMAP) and Retrotransposon-based insertion polymorphism (RBIP) marker systems based on Nana and Divo detected a high level of polymorphism and revealed the unique transpositional history of each element family. Phylogenetic trees based on a combination of SSAP and REMAP data show that markers for Nana and Divo are able to resolve different species lineages in the *Coffea* genus according to time and rate of transposition of each elements. Altogether our results indicate the usefulness of the retrotransposon-based marker system to study diversity in *Coffea*. In the future the use of several additional LTR-retrotransposons, harboring distinct transpositional activity histories, will be necessary to completely resolve the genetic diversity and relationships of the complex *Coffea* genus. This study is the first report of the development of a retrotransposon-based marker in *Coffea*.

An Updated Annotation Of NBS Domains In *Coffea arabica*

Alvaro Gaitan[1] , **Narmer Galeano**[1] , **Marcela Yepes**[2] , **Herb Aldwinckle**[2]

[1] Colombian National Coffee Research Center –CENICAFE, Federación Nacional de Cafeteros de Colombia - FEDERACAFE-Chinchiná, Caldas, Colombia
[2] Department of Plant Pathology and Plant Microbe Biology, Cornell University, Geneva, New York, USA

The majority of resistance genes isolated from plants encode a very conserved nucleotide binding domain (NBS). The domain is involved in protein-to-protein interactions that trigger metabolic pathways responsible for blocking pathogen development at the cellular level, among others. In *Coffea*, over 300 NBS

sequences have been identified by means of PCR amplifications, cDNA libraries and BAC end sequencing. Structural analyses were carried out using MEME to define the motif composition within the domain. In protein models obtained with MODELLER based on the human apoptosis gene APF-1, that contains the anchor points for ADP ligation, 95% of the amino acids displayed steric conformations allowed in Ramachandran diagrams. Grouping of potential families was supported by Hydrophobic Cluster Analysis (HCA) and 3D topology of the models. As a result, 10 molecular models are proposed for the 10 families reported in *Coffea*, all located within the superfamily P-loop. The *Coffea* NBS domain composition was compared with that of sequenced reference genomes including *Arabidopsis, Populus* and rice, indicating that it is unique between genera. SSCP (Single Strand Conformation Polymorphism), comparisons between *Coffea arabica* accessions and the species *C. eugenioides* and *C. canephora* showed particular diversification patterns for each family, and physical mapping probing a fingerprinted BAC library suggests local chromosome duplications of family members in the genome. Structural and comparative analyses of NBS domains provide support for the understanding of basic mechanisms of resistance gene function and divergence, and complement the current efforts on functional genomics and whole genome sequencing in the *Coffea* genus.

Opportunities And Challenges Of *Coffea arabica* Production And Germplasm Conservation In Its Center Of Origin-Ethiopia

GIRMA, A SENBETA[1] , **GETU G. BEKELE**[1] , **TADESSE W.M GOLE**[2]

[1] Senior Researcher, Jimma Rearch Center, PO Box 1669, JImma, Ethiopia
[2] Senior Researcher, Ethiopian Coffee Forest Forum, P. O. Box 28513 Addis Ababa, Ethiopia

Ethiopia is the primary center of origin and genetic diversity of *Coffea arabica* L., the most widely cultivated coffee species in the world. Ethiopia exports some of the world's finest specialty "coffees of origin" including Yirgacheffe, Sidamo, and Harrar, produced by small-scale subsistent farmers. Coffee is crucial to the Ethiopian economy as first agricultural export contributing to 10% of the country's GDP, and generating 40% of foreign currency. Ethiopia possesses about 99.8% of the genetic diversity of *C. arabica*. Internationally supported and coordinated conservation efforts are urgent in Ethiopia to prevent irreversible losses of invaluable biodiversity, as this highly endangered species is very vulnerable to the increasing settlement and land-use pressures on the mountain rainforests. At times, low international coffee prices for producers forced many coffee farmers to convert coffee lands to produce cereals and other cash crops like Khat (*Catha edulis*). Except for Ethiopia, the world coffee industry is entirely

based on a very narrow gene pool for cultivated *C. arabica* which makes the crop particularly susceptible to diseases and insect pests (coffee berry disease, leaf rust, wilt, berry borer and nematodes), as well as to environmental calamity due to climate change (drought, frost, etc.). Given increasingly high demand for best quality coffee by consumers, it is a principal priority for the coffee scientific community and stakeholders to develop an international consortium to conserve the unique *C. arabica* genetic diversity in its primary center of origin in order to sustain the future of the industry.

ABSTRACT PRESENTED AT THE SEQUENCING COMPLEX GENOMES WORKSHOP

Coffea canephora Genome Sequencing, A Tool For Comparative Genomics And Efficient Crop Improvement

Alexandre de Kochko[1] , **Victor, A. Albert**[2] , **Alan, C. Andrade**[3] , **Xavier Argout**[4] , **Benoit Bertrand**[4] , **Giovanni Giuliano**[5] , **Giorgio Graziosi**[6] , **Robert Henry**[7] , **J Jayarama**[8] , **Philippe Lashermes**[1] , **Ray Ming**[9] , **Chifumi Nagai**[10] , **Steve Rounsley**[11] , **David Sankoff**[12] , **Patrick Wincker**[13]

[1] IRD Centre de Montpellier BP 64501 34394 Montpellier cedex 5, France
[2] Department of Biological Sciences, University at Buffalo (SUNY), USA
[3] EMBRAPA, Final Av. W/5 Norte Brasília-DF 70770-917, Brasil
[4] CIRAD 34398 Montpellier Cedex 5, France
[5] ENEA, Post Box 026, via Anguillarese 301 - 00123 S.M. di Galeria, Roma, Italy
[6] University of Trieste. P.le Valmaura . 34100 Trieste, Italy
[7] QAAFI, The University of Queensland. St Lucia 4072 Australia
[8] CCRI, Coffee Board, Chikmagalur Dist., Karnataka State, India
[9] University of Illinois at Urbana-Champaign. Urbana, IL 61801 USA
[10] HARC PO Box100, Kunia HI 96759-0100 USA
[11] Dow Agrosciences LLC, Indianapolis, IN, USA
[12] University of Ottawa, 585 King Edward Avenue, K1N 6N5 Ottawa, Canada
[13] Genoscope, Institut de Génomique du CEA, CP5706 91057 Evry cedex France

Coffee is the most traded crop by Southern countries and is the main source of income for more than 75 millions small farmers all over the world. Despite its economical importance and the obvious necessity to improve the crop in order to respond to new environmental constraints and to the consumer demand for quality, the sequencing of its genome only started in late 2009. An international consortium was formed, led by Genoscope, to perform this task. *Coffea canephora*, a diploid cultivated species, was chosen, since *C. arabica* is tetraploid. Furthermore, IRD developed a double haploid plant because *C. canephora* is allogamous. Its genome size is about 695 Mb and whole genome sequencing is

being performed using NGS complemented by BAC ends coming from two BAC libraries covering in total 14.8 genome equivalents. SSR markers mined from these sequences are being mapped to establish a consensus genetic map based on the map kindly provided by Nestlé and ICCRI. Both Roche pyrosequencing (454) and Illumina technology are used to provide a 20x coverage by 454 and 50x by Illumina. Direct and paired end sequencing are underway, two, 8kb and 20kb insert libraries have been constructed. In addition to the publicly available EST, more transcriptome sequencing is also planned using 454 to facilitate the annotation. The *Coffea* genome will be one of the first Asterid genomes to be sequenced providing information on the proposed ancestral eudicot genome hexaploidization and for comparative genomics among angiosperms. It will also provide information to breeders for relating QTLs to genes.

ABSTRACTS OF POSTER PRESENTATIONS ON COFFEE

Characterization Of Prolyl Oligopeptidase Genes Differentially Expressed Between Two Coffee Cultivars

Ratnesh Singh[1,2] , **Beth Irikura**[3] , **Chifumi Nagai**[1] , **Henrik H. Albert**[4] , **Monto Kumagai**[5] , **Robert E. Paull**[3] , **Ming-Li Wang**[1,2]

[1] Hawaii Agriculture Research center, Waipahu, HI 96797, USA
[2] Molecular Biosciences and Bioengineering, University of Hawaii, Honolulu, HI 96822
[3] Tropical Plant and Soil Sciences, University of Hawaii, Honolulu, HI 96822
[4] PacificBasin Agriculture Research Centre USDA-ARS, Aiea, HI, 96701, USA
[5] Miller Associates, Davis, CA, 95616, USA

Coffee (*Coffea arabica* L.) is an economically important crop of Hawai´i. 'Tall Mokka' (MA2) and 'Typica' (KO34) are two coffee cultivars renowned for their high quality beans. 'Tall Mokka' differs from 'Typica' in having a bushier appearance and smaller aerial organs. Differential gene expression analysis of the two cultivars, using a potato microarray as a cross species platform coupled with qRT-PCR, identified higher expression of a prolyl oligopeptidase (POP) gene in the shoot tips of 'Tall Mokka'. POP, a ubiquitous gene found in many forms of life, encodes for an enzyme in animals known to cleave small proteins (≤30aa) such as neuro-peptides and it interacts with the IP3 pathway in Dictyostelium. Studies of POP in plants are rare, therefore the functions of it in plants are not known. Cloning and sequencing POP from 'Tall Mokka' and 'Typica' revealed three variants of the gene arbitrarily named CaPOP1, CaPOP2 and CaPOP3. Ectopic expression of CaPOP1::CaPOP1 in transgenic Arabidopsis yielded a bushy phenotype having significantly more secondary branches than the wild

type. The AtPOP promoter directed stronger GUS expression in the guard cells of the leaf epidermis, at the tip of the developing leaf than in the middle of the leaf blade, and in the vascular tissues of the root. Overall, this study has added new information regarding POP and suggested directions for future studies on POP in plants.

Displacement Of Coffee Species In New-Caledonia Promotes Interspecific Hybridization By Altering Their Flowering Phenology

Valerie Poncet[1] , **Celine Gomez**[1] , **Serge Hamon**[1] , **Perla Hamon**[1] , **Marc Despinoy**[2] , **Alexandre de Kochko**[1]

[1] IRD, UMR DIAPC, 911 avenue Agropolis, BP 64501, 34394 Montpellier CEDEX 5, France
[2] IRD,Unité ESPACE (S140), 101 Promenade Roger Laroque – Anse Vata, BP A5, 98848 Cedex Nouméa, New Caledonia

Human disturbance of natural habitats or species introductions could promote hybridization between previously isolated species. For example, temporal isolation by divergent floral phenologies could be modified leading to hybridization between related species. The three coffee species (Coffea arabica, C. canephora and to a much lesser extent C. liberica) are native from inter-tropical Africa. These species are naturally adapted to different environments, and some traits are differentiated. In particular, since flowering occurs after 5–6 days, 7 days and 8 days for C. liberica, C. canephora and C. arabica, respectively, no natural cross pollination takes place between these species. In New Caledonia, these three species have been introduced and succeeded to survive and hybridize without human intervention in abandoned plantations. This suggests a deregulation of their flowering phenologies and an important role of the environmental factors. Contrary to the African conditions, no complete dry season and dissimilarities in terms of rain distribution throughout the year are observed and likely involved in deregulation of each species' phenology, thus weakening the effectiveness of this barrier to hybridization. Our additional objectives were: (1) to assess the extent and nature of hybridization events in abandoned plantations. A reference population was analyzed using microsatellite markers; (2) to assess the impact of climatic parameters on each of the two stages of the flowering process (initiation and blossoming) ; (3) to evaluate, based on both biological and ecological expertise and data obtained at the reference sites, environmental factors favourable for the introduced species' adaptation and hybridization. Gomez C. et al. 2010 RSE, 114(11):2731-2744. Gomez C, et al. 2010 JAE, 47(1):85-95.

MOCCAdb - An Integrative Database For Functional, Comparative And

Diversity Studies In The Rubiaceae Family

Christine Tranchant-Dubreuil , Olga Plechakova , Romain Guyot , Perla Hamon , Claudine Campa , Alexandre de Kochko , Serge Hamon , Valerie Poncet

IRD, UMR DIAPC Centre IRD de Montpellier BP 64501 34394 Montpellier CEDEX 5 France

In the past few years, functional genomics information has been rapidly accumulating on Rubiaceae species and especially on those belonging to the Coffea genus. An increasing number of expressed sequence tag (EST) data and EST- or genomic-derived microsatellite markers have been generated, together with Conserved Ortholog Set (COS) markers. This considerably facilitates comparative genomics or genetic maps based studies through the common use of ortholog loci across different species. Similar genomic information is available for tomato or potato, members of the Solanaceae family. Since both Rubiaceae and Solanaceae belong to the Euasterids I, integration of information on genetic markers would be possible and should lead to more efficient analyses and discovery of key loci involved in important traits such as fruit development, quality or adaptation. Our goal was to develop a comprehensive web data source for integrated information on validated ortholog markers in Rubiaceae. MOCCAdb (http://moccadb.mpl.ird.fr/) is an interactive online database that manages annotated and/or mapped PCR-based markers defined from ESTs and genomic sequence in Rubiaceae (Plechakova et al., 2009). Marker information was retrieved from 11 published works, and completed with original data validated in the respective laboratories. Markers were checked for redundancy, in vitro tested for cross-amplification and diversity status in up to 38 Rubiaceae. Users can search the database for markers, sequences, maps or information on diversity through multi-option query forms. MoccaDB also includes bioinformatics tools (CMap and BLAST) and hyperlinks to related external data sources (eg : GenBank or SOL Genomic Network database).

Differential Proteomic Of *Coffea arabica* Submitted To Different Post-Harvest Processes

Kalynka G. Livramento[1] , Anderson C. José[2] , Luciano V. Paiva[3] , Flávio M. Borém[4] , José D. Alves[5]

[1] LCBM/Federal University of Lavras (UFLA), Lavras, MG, 37200-000, Brazil
[2] DCF/UFLA, Lavras, MG, 37200-000, Brazil
[3] DQI/UFLA, Lavras, MG, 37200-000, Brazil

[4] DEG/UFLA, Lavras, MG, 37200-000, Brazil
[5] DBI/UFLA, Lavras, MG, 37200-000, Brazil

Besides the genotype and cultivation conditions, it is known that the quality of the coffee is strongly related to the biochemical process that takes place inside of the seeds after harvesting of the grains. The objective of this work was to compare the profile of the proteins distinctly expressed in C. arabica grains submitted to four different treatments: pulped coffee, coffee dried on yard and in dryer at 60oC and natural coffee, submitted to the same drying conditions till they reach the water content of 11% (b.u). Proteomic analysis was accomplished by bi-dimensional electrophoresis. The gels stained with Commassie Blue G-250 were analyzed by the ImageMaster 2D Platinum 5.0 program (Amersham). The proteins with a expression difference were sequenced by MALDI ToF/ToF mass spectrometry. Protein points was showed a high similarity with globulin 11S, glyceraldehyde-3-phosphate dehydrogenase, late embryogenesis abundant protein, dehydrin DH1a, alpha galactosidase precursor and utp-glucose-1-phosphate uridylyltransferase. These proteins identified included many previously characterized stress-responsive proteins and others related to processes including storage protein, pentose and glucuronate interconversions, Starch and sucrose metabolism, Galactose metabolism, processing and degradation; and metabolism of energy, amino acids and hormones. The results of this study have shown that proteomics analysis was efficient in the coffee differentiate biochemically subjected to different post-harvest processes. Supported by CAPES, CNPq e FAPEMIG.

ABSTRACTS FROM OTHER CROPS THAT INCLUDED COMPARATIVE STUDIES WITH COFFEE

Pathway Networks For Cereals

Palitha Dharmawardhana[1] , **Liya Ren**[2] , **Marcela Monaco**[2] , **Jim Thomason**[2] , **Vindhya Amarasinghe**[1] , **Doreen Ware**[2,3] , **Pankaj Jaiswal**[1]

[1] Department. of Botany and Plant Pathology, 2082 Cordley Hall, Oregon State University, Corvallis, OR, 97331-2902, USA
[2] Cold Spring Harbor Laboratory, 1 Bungtown Rd, Cold Spring Harbor, NY, 11724, USA
[3] USDA-ARS NAA Plant, Soil & Nutrition Laboratory Research Unit, Cornell University, Ithaca, NY, 14853, USA

Gramene is a curated comparative plant genomics database (www.gramene.org). In order to understand the metabolome of cereal crop plants, we develop and

curate pathway databases for cereal species. The portal includes pathways for rice (RiceCyc) and *sorghum* (SorghumCyc) which are updated twice a year and the new databases for maize (MaizeCyc) and *Brachypodium* (BrachyCyc). In addition, the plant metabolic pathways module within Gramene mirrors several other species specific-pathway databases for plants such as *Arabidopsis* (AraCyc), *Medicago* (MedicCyc), and tomato, potato, pepper and **coffee** (SolCyc), as well as the reference pathway databases MetaCyc and PlantCyc. Comparative analysis feasible through this collaboration between PMN, SGN, MetaCyc and Gramene will be discussed. Having all the individual pathway databases allows user to extract inter- and intra-specific comparisons between pathways and, associated genes and to look for patterns of coexpression. Since the database holds information in a network format for genes, gene products (RNA and protein), compounds, enzymes, reactions and pathways, an integrated tool called the Omics Viewer allows users to overlay and visualize transcriptomic, proteomic, and metabolomic datasets with expressed values on pathway maps. We have also built an Omics Validator tool to validate user provided expression data files by mapping probe IDs from various microarray platforms to their respective gene IDs. The gene sets have cross-reference links to their respective genomes if available. Compounds and reactions have cross-reference links to resource libraries like KEGG, CAS, LIGAND, ChEBI and PubChem. Work is currently underway to include regulatory and signaling pathways for rice.

Fingerprinting The Asterids Species Using Subtracted Diversity Array Reveals Novel Species-Specific Sequences

Nitin Mantri[1] , **Alexandra Olarte**[1] , **Chun Guang Li**[2] , **Charlie Xue**[2] , **Edwin Pang**[1]

[1] School of Applied Sciences, Health Innovations Research Institute, RMIT University, Melbourne, 3000, Victoria, Australia
[2] Division of Chinese Medicine, Health Innovations Research Institute, RMIT University, Melbourne, 3000, Victoria, Australia

Accurate fingerprinting of medicinal plant species is crucial for quality control and safety of herbal medicine. To aid authentication of medicinal species from Asterids, one of the major plant clade, an Asterids-specific microarray (ASM) was constructed by subtracting DNA of 104 non-asterids species from DNA of 67 Asterids species. 283 Asterids-specific sequences were used on a microarray platform to fingerprint 25 species representing 20 families and 9 orders within asterid clade. All the 25 asterid species tested using this array generated unique hybridisation patterns allowing to discriminate between them. Three species (Cornus spp.; Gardenia jasminoides; Lonicera japonica) that were not used to make initial DNA pool for subtraction also hybridised to spots on the array and

produced unique fingerprints allowing differentiation. The spots that were most polymorphic from a PCA analysis were of two types. One set of spots hybridised to most of the species tested but produced a signal of different intensities allowing discrimination between species. Other set of spots specifically hybridised only to a particular species or a particular family. Sequence analysis revealed that the spots that hybridised to most of the species tested were mostly identical to different chloroplast genes used in 'bar-coding of plants'. Interestingly, many spots that were species- or family-specific had no significant matches in the NCBI database meaning these are novel species-specific sequences. Amongst those are spots that are specific to *Lonicera japonica* (honeysuckle), *Angelica archangelic* (wild celery), *Ilex paraguariensis* (yerba mate), *Leonurus cardiaca* (motherwort), *Achillea millefolium* (yarrow), and *Coffea arabica* (**coffee**).