International Coffee Genomics Network (ICGN)

Report 6th Coffee Genomics Workshop held at the XXI Plant and Animal Genome (PAG) Meeting San Diego, California January 12-16, 2013

Coffee Genomics Workshop Speakers

1. **France Denoeud,** CEA-Genoscope- France. The *Coffea canephora* genome.
2. **Roman Guyot**, IRD-France. LTR Retrotransposons in the Robusta coffee genome (*Coffea canephora*): Identification and charaterization of elements from BAC-end and 454 sequences
3. **Philippe Lashermes,** IRD-France. Contribution of subgenomes to the transcriptome and their intertwined regulation in the allopolyploid *Coffea arabica* L. grown in contrasted temperatures
4. **Luiz Filipe Protasio Pereira**, EMBRAPA-Brazil. RNA-Seq analysis and *de novo* transcriptome assembly of *Coffea arabica* and *Coffea eugenioides*
5. **Cecília A. F. Pinto-Maglio**, Instituto Agrónomico, Campinas –Brazil. FISH-mapping pachytene chromosomes of coffee (*Coffea* L.)
6. **Marco Cristancho** and **Huver Posada**, CENICAFE- Colombia. The genome of coffee leaf rust, the most devastating pathogen of coffee, and development of molecular markers for marker-assisted selection in coffee breeding programs.

Coffee Genomics Workshop at PAG

The Plant and Animal Genome (PAG) meeting is the largest international scientific conference reporting on animal and plant genomics developments in the world, this year with more than 2,800 participants from 65 countries. The XXII Plant & Animal Genome Conference will be held in San Diego, January 11-15, 2014. For those interested in participating in future meetings see http://www.intlpag.org.

Approximately 50 scientists participated in our 6th coffee genomics workshop held as part of the PAG Meeting in San Diego on January 15, 2012. The co-organizers of the workshop, Marcela Yepes (Cornell University, my11@cornell.edu), Philippe Lashermes (IRD-CIRAD, France, philippe.lashermes@ird.fr), and Rod Wing (University of Arizona) thank the speakers for their participation and contributions. Abstracts of workshop and poster presentations 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 7th Coffee Genomics Workshop that will be held January 12, 2014 as part of the XXII PAG meeting in San Diego, January 11-15, 2014. Please contact one of the organizers if interested in presenting a talk or poster, or with suggestions for new topics for workshop presentations or for round table discussion at the ICGN meeting. 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.

ICGN survey and collaboration with the International Coffee Organization

ICGN conducted a survey in 2012 to help us update our mailing list, identify future priority projects for the community as well as new leadership to help secure funding for new proposals. ICGN members were asked to help us contribute to this effort by completing and submitting the survey available at our www site (http://www.coffeegenome.org). Survey results will be posted shortly, if you have not submitted your responses please do so. We appreciate your feedback.

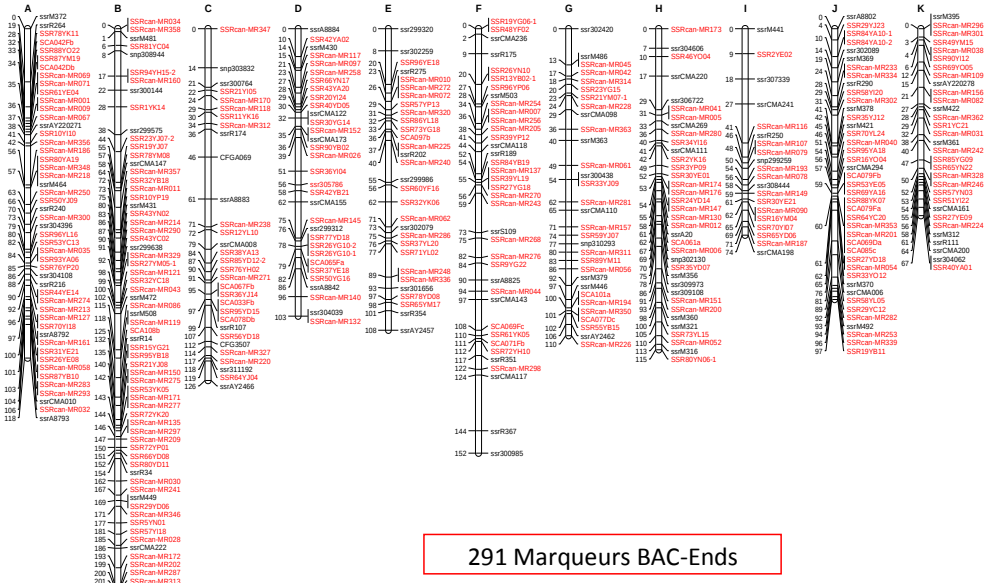
As the first *de novo* coffee genome reference become available (see report on the status of on going projects below), we would like to take advantage of the momentum to identify new priority projects of interest that ICGN can develop as a community to help mine the data generated and develop innovative tools and advanced resources in coffee genomics to address challenging issues for our community such as climate change adaptation and sustainability that could be accelerated with transforming genomic technologies and strategies. The African Coffee Research Network (ACRN) joined ICGN in 2011 as an institutional member, and its Director of Research and Development, Dr. Bayeta Bellachew helped us conduct the ICGN survey among ACNR members at several Coffee Research Institutions in Africa. We received through ACRN responses from scientists and scientific groups from the following countries: Ethiopia, Kenya, Rwanda, Uganda and Ghana with strong interest to work with ICGN on a global initiative to develop advanced genomic tools to speed up diversity characterization, enhanced utilization and conservation of *Coffea* germplasm in the context of climate change. In addition with support from the International Coffee Organization (ICO), ICO member countries have been contacted to discuss possible interest in developing a global initiative in collaboration with ICGN/ICO aiming at improving conservation and characterization of the world coffee gene pool for varietal development in a world of changing farming systems and climate. Other ICO member countries that have expressed strong interest in working on an ICGN/ICO collaborative proposal include, for Europe: France (IRD-CIRAD); for Latin America: Brazil, Colombia, Guatemala, Costa Rica, Mexico; for Africa: Cote D'Ivoire, Ethiopia, Kenya, Malawi, as well as the Inter-African Coffee Organization; and for Asia: India and Vietnam.

ICGN is grateful for the invitation by the new ICO Executive Director Dr. Robeiro Oliveira Silva to participate as an observer in the ICO Council meetings in 2013, and we are looking forward to working closely with ICO officials on the preparation and submission of a first ICGN/ICO proposal, and to explore potential sources of finance for such joint initiative. Support from ICO will be key for ICGN to secure future funding for diversity conservation efforts in *Coffea* with a broader funding base. The importance of cooperation on innovative coffee research for climate change adaption is key for future sustainability of the world coffee industry and working with ICO will strengthen ICGN efforts to secure funding for such initiative. ICO could help ICGN as a platform to help us share information for the community, help us fund outreach training opportunities in marker-assisted breeding, genomics and bioinformatics, as well as to secure funding for diversity studies and testing of germplasm and crosses generated to benefit breeding programs in multi-location trials for adaptation to climate change. ICO involvement will help us enhance collaborations among coffee research organization in different countries. ICO could help us support regional meetings with researchers interested in coffee genomics to include scientists from developing countries. Progress on coffee genomics research and the potential of research results for coffee improvement can be discussed, targeting the priority traits for different regions as well as for the coffee industry. Capacity building in developing countries to participate in coffee genomic research can be supported through ICO networking to help us secure funding for those efforts.

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 Nestlé R&D Centre 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. DNA from the two parental clones and the segregating progeny were distributed to several ICGN members (on request). First, a high-density genetic map was constructed including 1481 loci covering 1400cM markers, with a third

of the SSR markers derived from BAC end sequences (see Figure below). The first set of markers mapped included: 360 RFLPs, >890 SSRs, and 213 SNPs, and were mined from genomic or EST libraries from different institutes (IRD, CIRAD, Trieste University, Cornell University, CENICAFE, and Nestlé).



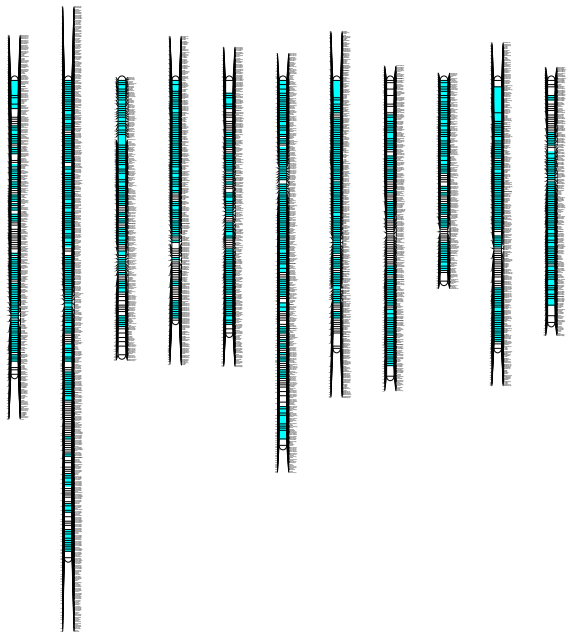
As a second approach, Restriction Associated DNA sequencing (RADseq), which enables synchronous SNP marker discovery and genotyping using massively parallel sequencing, was used. The RAD libraries were made from digestion of DNA using two restriction enzymes, *Nsi*I (6 base cutter) and *Mse*I (4 base cutter). The fragments (150 - 500 bp) were selected to ligate to two adaptors, and one of them with tag for each progeny. Equal amount of amplicons from each individual was pooled to make Illumina RNAseq libraries with individual tags for each library. Co-segregating markers within 50 kb region (< 1 cM) based on the aligned template scaffold were sorted as bin. One marker from each bin was selected for mapping. The linkage analysis and map construction were performed using JoinMap software version 4.1 using LOD threshold of 5 and Kosambi's function to calculate genetic distance between two loci. The Robusta consensus genetic map was built using the F2 segregating loci as anchor markers in order to merge the two homologous parental linkage groups. Using RAD sequence data from the segregating population previously selected 1747 RADseq markers were added. The final high density Robusta map comprises 3230 loci, genetic size 1471 cM (1cM ~ 500 Kb), with an average density close to one marker every 220 kb. The F1 high density genetic map will facilitate comparative genomic studies based on synteny and provides the opportunity for anchoring and ordering the numerous scaffolds arising from the *Coffea canephora* genome sequencing (see report below). So far, the DNA sequences (scaffolds) anchored are covering approximately 75% of the genetic map (1023 cM). On going mapping efforts are focusing on the identification and mapping of SSRs from the *C. canephora* sequence scaffolds that are not or are insufficiently anchored. See table and figure below summarizing the markers mapped so far on the high density *C. canephora* map.

Table 1: International reference genetic *Coffea canephora* map characteristics. The eleven linkage groups (A to K) were described according to their size (cM), number of markers, density and origin of the markers (RADs, RFLPs, SSRs and SNPs). The estimation of the sequencing scaffolds coverage is also given.

Linkage group	Size (cM)	Nb. Markers	Density	RADs	RFLPs,SSRs, SNPs	Scaffold coverage (cM)	Scaffold coverage (%)
A	136	320	0.43	157	163	86	63
B	221	520	0.43	304	216	152	69
C	127	235	0.54	136	97	90	71
D	111	274	0.41	149	125	81	73
E	117	266	0.44	140	126	84	72
F	169	349	0.49	168	181	123	73
G	124	305	0.41	156	149	98	79
H	137	271	0.51	168	103	76	55
I	93	180	0.52	98	82	68	73
J	124	286	0.44	156	130	88	71
K	112	224	0.50	115	109	77	69
Total	1471	3230	0.46	1747	1481	1023	70

Both the high-density genetic map and the marker information will be freely available on a dedicated web-site once the construction of the map is completed. Please send information or comments to Dominique Crouzillat, Nestlé (dominique.crouzillat@rdto.nestle.com), Philippe Lashermes (philippe.lashermes@ird.fr), or Ray Ming (rming@life.illinois.edu).

The international *C. canephora* high density map is a highly valuable resource for different applications including transposition to other mapping populations, as genetic framework that can be used for various QTL studies, as well as genome structure comparisons. RAD sequencing is a powerful strategy for genotyping in coffee to provide access to high-throughput SNP detection.



Update status of the *Coffea canephora* genome sequencing

With funding from the Agence Nationale de la Recherche (ANR), France, several Institutes (Genoscope-CEA, IRD and CIRAD) combined their scientific resources and expertise to sequence, assemble, and annotate the entire genome of *C. canephora*. Additional partners include several ICGN members (EMBRAPA/Brazil, ENEA/Italy, University of Trieste/Italy, University of Queensland/Australia, CCRI/India, University of Illinois, Urbana/USA, Hawaii Agriculture Research Center HARC/USA, SUNY Buffalo/USA, University of Ottawa/Canada). A community effort for genome annotation is on going. The *C. canephora* genome consists of 11 chromosomes, is about 710 Mb in size, and was sequenced *de novo* with deep coverage using different sequencing platforms. Genoscope lead the sequencing and assembly of the *C. canephora* genome. 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 2011 coffee genomics workshop, and this year France Denoeud from Genoscope presented in our coffee genomics workshop an update on the first genome assembly, see copy of the abstract in the appendix.

C. canephora is one of the ancestral progenitors of the widely cultivated, *C. arabica*, a recent allotetraploid species formed from the merger of the diploid species *C. canephora* and *C. eugenioides*. The accession DH200-94, a doubled haploid genotype was selected for sequencing because of its homozygous nature to facilitate genome assembly. *De novo* genome sequencing with deep coverage was performed using both 454 Roche and Illumina technologies. Direct whole genome shotgun (WGS) sequencing and paired-end sequencing of large insert libraries 8kb and 20 kb insert libraries was conducted. Furthermore, clones from two *C. canephora* BAC libraries were BAC-end sequenced using Sanger technology. The overall data generated was as follows:

Roche/454:

WGS 454 sequencing : 28.9X coverage (assuming a genome size of 710 Mb) including:
23X single end 454 reads
– Reads single end: 14.83X , Mean size: 359 bp
– Long reads single end: 8.24X , Mean size : 462 bp
5.8X Paired-end sequencing of long insert libraries (8 and 20 Kb): 5.8X (2.2X for 8kb, 3.6X for 20kb), Mean Size: 252 bp.

Sanger:

Two BAC libraries (*Hind* III and *Bst*YI) were constructed in collaboration with Rod Wing at Arizona Genomics Institute. The BAC libraries have 73,728 clones (>11X coverage).
Sanger BAC-end sequencing: 131,412 BES were generated (73,728 BAC clones x 2 ends 5' and 3'): 0,27X
Mean BAC insert size : 135 Kb, range: 63,2Kb < insert < 253,6 Kb

Illumina sequencing was done at deep coverage (~70X) to correct 454 sequencing errors.

Single reads coverage 7.3X: read size 76 bp (4.8X) and read size 150 bp (2.5X)
Paired end reads coverage 62.4X: read size 76 bp (42.4X) and read size 108 bp (20X)

Assembler used Newbler and assembly statistics for the first assembly (as of Jan 2013) (data from Genoscope):

No. of Scaffolds 13,345
Size: 569 Mb (80% estimated genome size of 710 Mb)
Coverture 28.87 X (454/Sanger) and Illumina 69.7 X
N50: 1260 Kb (108 scaffolds)
N80: 65 kb (635 scaf)
Largest scaffold: 9.0 Mb

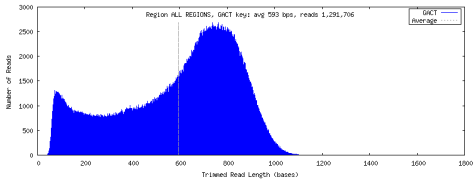
Automatic annotation by reconciliating proteic and EST hits, RNA-Seq data integrated into G-Mo-R-Se models, and *ab initio* predictions, was performed and provided 25,574 genes (with an average of 5.1 exons per gene). This relatively low gene number compared to other plant lineages reflects the fact that the *Coffea* lineage was less frequently subject to whole genome duplications than other sequenced dicotyledon lineages. Its slow rate of genome evolution makes it a good model for paleogenomics in Asterids.

Update status of the *Coffea eugenioides* genome sequencing

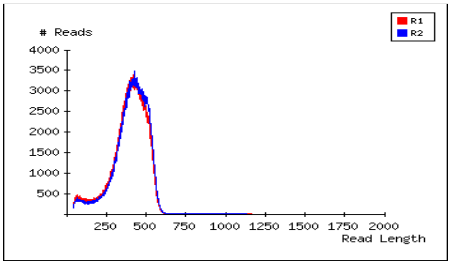
This project is funded by the InterAmerican Development Bank (FONTAGRO/SECCI), and sequencing was started in the Fall of 2012. The project is being developed collaboratively by the Colombian National Coffee Research Center (CENICAFE), and Cornell University. Funding for this project was secured jointly through a proposal prepared and submitted by Cornell University and CENICAFE.

We are mimicking the strategy used for the *C. canephora* sequencing to generate a high quality reference assembly for *C. eugenioides* using mixed next generation sequencing platforms: Roche 454 FLX+ and Illumina HiSeq 2500. We have sequenced a whole genome shotgun (WGS) library (fragment size >1,100 bases and <2,000 bases) using 454 FLX+ single end reads with mode length of 763 bases to generate a total of 6,082,341,937 bases for an estimated coverage of 9.2X for the *C. eugenioides* genome (~estimated genome size of 660 Mb). See below quality control results of a typical 454 FLX+run:

Region	Passed Filter	Total Bases	Reads ≥ 1000 bases (% Passed Filter)	Read Length (Bases) Mean	Median	Mode	Min	Max
1	634,065	373,033,536	4,712 (0.74%)	588.3	653	763	40	1,735
2	657,641	392,830,440	3,723 (0.57%)	597.3	660	782	40	1,692
ALL REGIONS	1,291,706	765,863,976	8,435 (0.65%)	592.9	657	763	40	1,735



We have completed sequencing of twelve 20 Kb long insert libraries (3.1X coverage) using paired end sequencing and Roche 454 FLX Titanium. Below is a graph of the paired end statistics profile for a run: Read length distribution of high quality reads. R1 = Coffee 20kb 1-1, R2 = Coffee 20kb 1-1. Linker positive displays statistics of reads with paired end linker sequence 71.47% and 73.62%. Linker Negative refers to reads with no paired end linker sequence 28.53% and 26.38%.



We are currently constructing 40 Kb long insert libraries for paired end sequencing using 454 FLX Titanium, and will perform Illumina deep sequencing with different insert size libraries prior to initiate genome assembly. 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 frames for sequencing and assembly of *C. arabica*, the major cultivated coffee species worldwide.

By sequencing the coffee genome, we are building 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 will be 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 full benefit from 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.

Acknowledgements:
ICGN is particularly grateful to all our workshops speakers who kindly accepted our invitation to participate as speakers in our coffee genomics workshop. Abstracts of their presentations are enclosed below.

Appendix

Abstracts 6th Coffee Genomics Workshop 2013

Workshop Co-Organizers:

Marcela Yepes, Cornell University (my11@cornell.edu)
Philippe Lashermes, L'Institut de Recherche pour le Développement (IRD), France (philippe.lashermes@ ird.fr)
Rod Wing, University of Arizona (rwing@Ag.arizona.edu)
(Program and abstracts also posted at:
https://pag.confex.com/pag/xxi/webprogrampreliminary/Session1503.html

The *Coffea canephora* genome

France Denoeud CEA-Genoscope, EVRY CEDEX, France

We sequenced the genome of *Coffea canephora* with the following strategy: 29X of 454 reads (23X single end, 6X paired reads of 8 kb and 20 kb) and 0.2X of BAC ends Sanger sequences were assembled using Newbler. 70X of Solexa reads (7.3X single end, 62.4X pairedend) were then used for assembly correction and gap closure. The final assembly contains 569 Mb, with an N50 (scaffolds) of 1260 kb. We performed an automatic annotation by reconciliating proteic and EST hits, RNA-Seq data integrated into G-Mo-R-Se models, and *ab initio* predictions, which provided 25574 genes (with an average of 5.1 exons per gene). This relatively low gene number compared to other plant lineages reflects the fact that the *Coffea* lineage was less frequently subject to whole genome duplications than other sequenced dicotyledon lineages. Its slow rate of genome evolution makes it a good model for paleogenomics in Asterids. We are presenting the *Coffea canephora* genome landscape and comparative genome structure analysis.

LTR Retrotransposons in the Robusta coffee genome (*Coffea canephora*): Identification and charaterization of elements from BAC-end and 454 sequences

Elaine Silva Dias[1,2], Clemence Hatt[1], Christine Dubreuil-Tranchant[1], Alexis Dereeper[1], Philippe Lashermes[1], Serge Hamon[1], Michel Rigoreau[3], Dominique Crouzillat[3], Perla Hamon[1], Claudia Carareto[2], Alexandre de Kochko[1] and **Romain Guyot**[1], (1)Institut de Recherche pour le Développement, Montpellier

cedex 5, France, (2)Laboratory of Molecular Evolution UNESP, São José do Rio Preto, SP, Brazil, (3)Centre R&D Nestlé Tours, France

Coffee is one of the most important international trade commodities and is ranked as the second most valuable primary commodity exported by southern countries. Two species are mainly used in commercial production: *Coffea arabica*, known as Arabica and *Coffea canephora*, a perennial diploid species known as Robusta. Recently, 131,412 BAC-end (BES) and 106,459 454 sequences were generated from a *C. canephora* Double Haploid accession (DH200-94), which corresponds to almost 19 % of the estimated *C. canephora* genome size. Here we present the identification and the characterization of LTR Retrotransposons in the *C. canephora* genome based on the analysis of BES and 454 sequences. To reconstruct the structure of LTR Retrotransposons, we used the Assisted Automated Assembler of Repeat Families algorithm (DeBarry et al., 2008) to assemble sequences into 37 different LTR-RTs contigs. The contig sizes range from 11,130 to 3,070 bp and nine of them were annotated as complete LTR-RTs elements. Based on a RT domain classification, we show that all LTR-RTs contigs belong to well defined LTR-RTs families. Among the 37 LTR-RTs contigs, 27 and 10 belong respectively to the *Gypsy* and *Copia* superfamilies. Twenty-three LTR-RTs contigs were found highly conserved in *C. canephora* EST sequences, and eight were found transcriptionally active in *C. canephora* leaves and/or fruits, suggesting that these LTR-RTs are potentially expressed. One of the LTR-RTs contigs assembled here and called *COPIA25* was characterized in details. *COPIA25* belongs to the *Copia Tork4* family of LTR-retrotransposons. It was found expressed by PCR amplification on *C. canephora* cDNA libraries. To investigate the origin of *COPIA25* in coffee we retrieved 24 highly similar sequences from 18 model genomes and we analyzed their phylogenetic relationships. Surprisingly, elements from potato and banana clustered with *COPIA25*. A detailed analysis confirmed an outstanding nucleotide (75%) and amino acid conservations (82%) between *COPIA25*and the potato/banana elements. Considering the species divergence between coffee and banana (> 150 My) and coffee and potato (83-89 My), such remarkable conservation may suggest events of TE horizontal transfers. Additional characterization of *COPIA25* will produce new insight into the impact of this family into the *Coffea* genome evolution and diversity.

Contribution of subgenomes to the transcriptome and their intertwined regulation in the allopolyploid *Coffea arabica*, L. grown in contrasted temperatures

Marie-Christine Combes , Alexis Dereeper, Benoit Bertrand , **Philippe Lashermes**
Institut de Recherche pour le Développement, CIRAD, Montpellier cedex 5, France

Polyploidy occurs throughout the evolutionary history of many plants, giving rise to novel phenotypes and leading to ecological diversification and colonization of new niches. *Coffea arabica* a recent allopolyploid between two low divergent diploid species, *Coffea eugenioides* and *Coffea canephora*, can be grown in regions with marked variations in thermal amplitude while the parental species are less adapted to temperature variations. To assess the contribution of subgenomes to the *C. arabica* transcriptome and its variation in relation to the adaptation to variable culture conditions, the transcriptome of leaves of *C. arabica* plants grown in two contrasted thermal regimes were examined using RNA sequencing (RNA-seq) approach. Then the relative homoeologous gene expression was compared to the relative expression between the modern-day diploid progenitor species. Whatever the growing conditions, 65% of the studied genes showed equivalent level of homoeologous gene expression, and for a large majority (92%), the relative homoeologous gene expression between both growing conditions varied less than 10 %. In *C. arabica* growing conditions do not disrupt the relative homoeologous gene expression and the overall gene expression appears to be regulated by intertwined mechanisms. Unlike others allopolyploids analyzed previously, *C. arabica* showed no preferentially expressed subgenome illustrating the interest of considering the age of the allopolyploidization event and the evolutionary divergence of progenitor species when studying allopolyploidy.

RNA-Seq analysis and *de novo* transcriptome assembly of *Coffea arabica* and *Coffea eugenioides*

Priscila Mary Yuyama[123], Marcelo Falsarella Carazzolle[4], Osvaldo Reis Júnior[4], Jaime Hikaru Mishima[4], Suzana Tiemi Ivamoto[23], Douglas Silva Domingues[2], Pierre Charmetant[12], Thierry Leroy[1], **Luiz Filipe Protasio Pereira**[5]1. CIRAD.

Avenue Agropolis. 34398 Montpellier Cedex 5. France.2. IAPAR, Laboratório de Biotecnologia Vegetal. Rodovia Celso Garcia Cid Km 375. 86047-902, Londrina - PR. Brazil.3. Universidade Estadual de Londrina, Centro de Ciências Biológicas, Departamento de Biologia Geral. Campus Universitário, 86055-900, Londrina – PR. Brazil. Universidade Estadual de Campinas, Instituto de Biologia, Departamento de Genética e Evolução. *Laboratório de Genômica e Proteômica*. Barão Geraldo, 13083-970, Campinas – SP. Brazil. Embrapa Café, 70770-901, Brasília – DF. Brazil.

Coffee is one of the most important agricultural commodites in the world. *Coffea arabica*, an allotetraploid from a recent hybridization of two diploid species (*C. canephora* and *C.eugenioides*) is responsible for 70% of world commercial production. Recent studies have been done to analyse transcriptome data of *Coffea* spp. in order to improve our knowledge in genetics and gene expression of those species. However, there is few data on coffee fruit transcriptome. In this work, RNA-Seq of two organs (whole fruit – cherry stage and mature leaves) from allotetraploid *C. arabica* cv. Mundo Novo and its diploid ancestor *C.eugenioides*, were done using Illumina HiSeq 2000 (*100 bp single-end sequences).* We report the generation of near 10 million reads for *C.arabica* (4,747,049 of fruits and 6,156,750 reads of leaves) and 7,5 million reads for *C.eugenioides* (3,688,364 of fruits and 3,835,373 reads of leaves). *De novo* assembly and digital gene expression analyses was performed to generate a collection of 35,462 contigs from *C. arabica* cv. Mundo Novo (average length of 691 bp) and 36,935 contigs of *C. eugenioides* (average length of 701 bp). Funnctional annotation were performed by sequence comparison with public databases (NCBI-nr, Swiss-prot and PlantCyc). Our results showed the presence of prevalent genes and species-specific either in *C. arabica* or *C. eugenioides* that could explain particular characteristics of these two species. Characterization of *Coffea spp*. transcriptome provides an effective tool for better understanding for differential gene expression in *Coffea* spp., providing important clues of *C.arabica* gene evolution, as well as valuable information for coffee breeding.

FISH-mapping pachytene chromosomes of coffee (*Coffea* L.)

Ana Amélia Sanchez Iacia , and **Cecília A. F. Pinto-Maglio**, Instituto Agronomico, Campinas, Brazil

Fluorescent *in situ* hybridization (FISH) provides the most direct way of physically mapping DNA sequences on chromosomes. FISH-mapping to meiotic chromosomes at pachytene is an important tool in plant cytogenetic research as it provides good resolution measurements of physical distances. We have been able to overcome the most difficulties to apply a specific method for FISH on pachytene chromosomes of coffee plants (*Coffea* L.). Meiotic chromosomes observed at pachytene stage in pollen mother cells (PMCs) of coffee are more suitable for FISH physical mapping because they are about 30 times longer than their somatic counterparts. They also provide additional cytological markers like a pattern of distal euchromatin and pericentromeric heterochromatin regions. For the first time we observed hybridization signals on coffee pachytene chromosomes by FISH technique using a repetitive sequence as probe. The method involves some procedures for obtain suitable pachytene chromosomes. Slides with 80% of the PMCs with hybridization signals on bivalents were obtained. The method developed showed no genotype dependency as we registered the presence of hybridization signals in different coffee plants. The use of this method expands the possibilities for high definition physical mapping of coffee chromosomes.

The genome of coffee leaf rust, the most devastating pathogen of coffee, and development of molecular markers for marker-assisted selection in coffee breeding programs

Marco Cristancho[1], **Huver Posada[1]**, Álvaro Gaitán[1], William Giraldo[1], David Botero[1,2], Javier Tabima[1,2], Diana Ortiz[1], Alejandro Peralta[1], Silvia Restrepo[2], Diego Riaño[2], Fernando García[1], Camilo López[3], Marcela Yepes[4], and Herb Aldwinckle[4] [1]Colombian National Coffee Research Center – CENICAFE, Chinchiná, Caldas, Colombia; [2]Universidad de los Andes, Bogotá, Colombia; [3]Universidad Nacional de Colombia, Bogotá, Colombia; [4]Cornell University, Geneva, NY, USA

Coffee leaf rust, caused by the fungus *Hemileia vastatrix*, is the most damaging disease of coffee worldwide. The pathogen has caused yield reductions of up to 30% in susceptible varieties of *Coffea arabica* in recent epidemics associated with climate change in Colombia and several

Central American countries. We have sequenced *de novo* the genome and transcriptome of *H. vastatrix* race II using next-generation sequencing platforms: 454 Roche and Illumina. The *H. vastatrix* genome is over 250Mb, one of the largest fungal genomes to date. The assembled genome has 92% coverage, a GC content of 32%, and a very large proportion of repeated sequences; over 74% of the *H. vastatrix* genome is composed of repeats, with prevalence of LTR retro-transposons. We predicted a set of 14,445 proteins encoded by the coffee rust genome, and 44 secreted proteins that might be involved in pathogenesis. Three transcriptome assemblies from different isolates/races were obtained with 44,297, 55,791 and 64,752 contigs with average contig size of 675, 716 and 828 bp. Most *H. vastatrix* race II transcripts (84%) were mapped to the assembled genome. *H. vastatrix* isolate/race diversity at the transcriptome level is significant, since they only share 76-78% of expressed sequences. Only 36% to 44% *H. vastatrix* transcriptome sequences had homology with *Puccinia* or *Melampsora* proteins showing the high level of genetic diversity present in rust fungi. We have also developed molecular markers for *H. vastatrix* that are being used for genetic diversity studies.

In addition, we have developed molecular markers for host derived resistance against this pathogen that are linked to the introgression of *C. canephora* (rust resistant genotype*)* in breeding F_5 lines of *C. arabica*, particularly in the resistant components of the Castillo variety of Colombia. These markers have high potential to accelerate development of rust resistant varieties through marker-assisted selection in coffee breeding programs.

Note: This abstract will have two presenters: Marco Cristancho and Huver Posada.

Abstracts on coffee presented at PAG poster session:

Coffee transcriptome analysis using RNA sequencing: Comparison of coffees grown at different altitudes

Juan F. Medrano[1], Ricardo Koyner [2], Alma Islas-Trejo[1], Gonzalo Rincon[1], [1]University of California-Davis, Davis, CA, [2]Kotowa Farms, Boquete, Panamá

The objective of this study was to develop a transcriptome analytical framework of coffee beans to use as a reference to annotate and examine gene expression differences between coffee varieties and coffees grown at different altitudes. Coffee berries at different stages of maturity (unripe, half-ripe, cherry) were sampled from four plants (*C. Arabica-Geisha*) at two altitudes (1300m and 1650m) at Kotowa Farms, Boquete, Panama. It is known that the coffees grown at higher altitudes acquire unique flavor characteristics, whose molecular differences would be economically important to identify. Fresh bean slices were stored in RNAlater and ground in liquid nitrogen to extract RNA using Plant-RNA Reagent (Invitrogen). Bar-coded RNAseq libraries were sequenced using Illumina HiSeq. From 8 samples, 217 million 100bp single reads were used to perform a de-novo assembly using CLC Genomics software. A total of 119,678 contigs were generated (average contig size 527bp, N50 881bp and max-size 15,090bp). These contigs were used as reference to perform RNAseq analysis. Using Blast2Go software 21,000 contigs (17.5%) were annotated with known protein information. A t-test identified significant differentially expressed values (RPKM $P<0.01$) and fold changes greater than 2 in 206 contigs between cherry beans at 1300m and 1650m. The main GO terms for the 206 contigs corresponded to: ATP binding, nitrogen compound metabolic process, negative regulation of programmed cell death, anatomical structure development. This work contributes to the annotation of the coffee genome for the examination of differences between varieties and to study the effects of

environmental variables on economically important phenotypes.

Transcriptional changes associated to coffee resistance responses to the root-knot nematode *Meloidogyne incognita*

Erika V. S. Albuquerque[1], Joseane Padilha[1], Sergio Alencar[1], Georgios Joannis Pappas[1], Anne-Sophie Petitot[2], Maria Fatima Grossi de Sa[1], Diana Fernandez[2], [1]Embrapa, Brasília, Brazil, [2]Institut de Recherche pour le Développement, Montpellier, France

Root-knot nematodes (*Meloidogyne* spp.) are major pests damaging the coffee culture (*Coffea arabica*). To gain knowledge about the molecular mechanisms of plant immunity in coffee, we investigated the transcriptomic changes associated with the hypersensitive response (HR) to *Meloidogyne incognita* in resistant *C. arabica* 'UFV 408-28' plants. The gene expression profiles of 88 selected genes were analyzed by real time RT-qPCR at 4, 5 and 6 days post inoculation (dpi), corresponding to the onset of HR, in three independent experiments. First, we performed statistical analysis to verify the best model to fit each gene for inoculation, genotype, and the interaction inoculation x genotype effects. Coffee molecular responses were gradually increasing with infection time, with altered expression profiles observed for 4 (4 dpi), 12 (5 dpi) and 21 (6 dpi) genes. Interestingly, a WRKY transcription factor gene and a P450 gene were activated in incompatible interaction and suppressed in compatible interaction. Second, we compared the significant qPCR expression results to deep sequencing (Illumina) data obtained from resistant roots inoculated with nematodes or not, at 6 dpi. We tested the Trinity method for *de novo* assembly of Illumina transcripts without a coffee reference genome. An average of 3400 Illumina contigs were differentially expressed during coffee resistance responses to *M. incognita*. In particular, genes linked to the phenyl-propanoid pathway and salicylic acid biosynthesis were activated. Transcriptomic data obtained by both approaches were consistent, and some specific novel cDNA sequences may be used to unveil the resistance mechanisms acting in *C. arabica.*

Transcriptional and promoter analysis of CaDUR3, a *Coffea arabica* urea transporter gene

Douglas Domingues[1], Tiago B. dos Santos[1], Anderson R. Meda[1], André Y. Kashiwabara[2], Sandra M. B. Cação[1], Luiz G. E. Vieira[3], Luiz F. P. Pereira[1], [1]Plant Biotechnology Laboratory, IAPAR, Londrina, Brazil; [2]Universidade Tecnologica Federal do Paraná, Cornélio Procópio, Brazil; [3]UNOESTE, Presidente Prudente, Brazil

Nitrogen is one of the most expensive nutrients to supply and commercial fertilizers represent one of the major costs in coffee production. Urea is the major nitrogen form supplied as fertilizer in agriculture. Despite the importance of efficient use of urea-containing nitrogen fertilizers by crops, membrane transporters that might contribute to the uptake and transport of urea in plants were characterized only in model systems. Coffee is one of the most important agricultural commodities in the world, but molecular responses to environmental factors that may affect coffee production are still poorly understood. In this work, we evaluated the transcriptional pattern of *Coffea arabica* DUR3 (*CaDUR3*), a membrane transporter that plays an effective role in improvement of urea acquisition in rice and *Arabidopsis. CaDUR3* was mostly expressed in leaves and secondary roots, while little to no expression were observed in flower buds, flowers, plagiotropic shoots and pericarp of mature fruits. Transcriptional analysis by qPCR indicated that this gene was upregulated under nitrogen starvation, water deficit, high salt and heat stress. To determine the molecular mechanisms underpinning the differential expression of this gene, the upstream region of the gene was analysed in silico and putative binding sites for Dof, MYB and WRKY transcription factors were identified in *Arabidopsis* and *Coffea* orthologs. Further studies in gene structure and functional analysis in heterologous systems may confirm if *CaDUR3* is an active urea transporter that plays a significant role in urea acquisition and utilization in coffee plants. Finnancial support: Cnpq, FINEP and Consorcio Pesquisa Café.

Plant Pathways & Gene Expression Analysis in *Gramene*

Sushama Naithani and Pankaj Jaiswal , Dept. of Botany & Plant Pathology, Oregon State University, Corvallis, OR

The pathways section in the Gramene harbors species-specific metabolic pathway databases for rice (RiceCyc), maize (MaizeCyc), *Bracypodium* (BrachyCyc) and sorghum (SorghumCyc). These databases were constructed and curated by Gramene. We also mirror the reference databases; EcoCyc, MetaCyc and PlantCyc, in addition to pathway databases of *Arabidopsis (AraCyc)*, tomato (LycoCyc), potato (PotatoCyc), **coffee (CoffeaCyc),** and *Medicago (MedicCyc)* developed by our collaborators. Researchers can search for genes, proteins, enzymes, reactions, metabolites, biochemical, and metabolic and transport pathways in any of these databases and compare two or more species. Furthermore, Omics-viewer tool in the pathway databases allows users to upload and analyze their high-throughput expression data (from microarray, proteomics, metabolomics etc.) and generate cellular overview of pathways. The plant genome browser allows users to view and search genome information from the chromosome to a base-pair level. It also allows researchers to upload and analyze their data such as those from transcriptome assembly, gene expression and SNP analysis. In the hands-on session, we will use publicly available RNA-Seq gene expression and SNP data to illustrate utility of these tools. Researchers are welcome to bring their own datasets for practice.

SGN: Solanaceae Genomes, Tools and the Breeder's toolbox

Isaak Y. Teckle and Lukas Mueller, Boyce Thompson Institute for Plant Research, Ithaca, NY

The SOL Genomics Network (SGN, solgenomics.net) is a repository for whole genome sequences, large scale genotype, genetic map, expression, and phenotype data of *Solanaceae* and closely related species. It provides access to data using in-house and externally developed open source annotation, analysis and visualization tools. Data is highly integrated and cross-referenced to allow discovery of knowledge. Crops with rich datasets at SGN include tomato, potato, pepper, tobacco, **coffee**, and petunia. Biologists and breeders interested in utilizing the data and tools available at SGN will benefit from attending the workshop. Bioinformaticians interested in exchanging experiences, ideas and co-developing tools are also welcome.

Upcoming Meetings of interest for the ICGN community

- 50th Anniversary International Coffee Organization (ICO), celebrating 50 yrs of International Cooperation on Coffee. Sep 9-13, 2013. Belo Horizonte, Minas Gerais, Brazil. http://www.ico.org/ico_meetings.asp?section=Meetings/Documents http://www.ico.org/presents/1213/march-brazil.pdf
- 7th Coffee Genomics Workshop at XXII Plant and Animal Genome (PAG) Meeting, San Diego, California, January 11-15, 2014 http://www.intlpag.org/
- 10th Solanaceae Genomics SOL meeting, Beijing, China, October 13-17, 2013. http://www.sol2013.org/
- 25th ASIC International Conference on Coffee Science, Colombia, 2014