MINUTES OF ICGN MEETING HELD AT THE PLANT AND ANIMAL GENOME CONFERENCE (PAG), SAN DIEGO, CALIFORNIA JANUARY 12, 2009

Participants (14)

Nestor Osorio	Executive Director International Coffee Organization
Patrick Wincker	Genoscope, France Head of Sequencing/Coordinator Eukaryote Annotation and Analysis
Chrispine O. Omondi	Coffee Research Foundation, Kenya
Elijah K. Gichuru	Coffee Research Foundation, Kenya
Carmenza Góngora	CENICAFE, Colombia
Felipe Rodrigues da Silva	EMBRAPA, Brazil
Alexandre de Kochko	IRD, France
Perla Hamon	IRD, France
Philippe Lashermes	IRD/CIRAD, France
Dominique Crouzillat	Nestlé, France
Giorgio Graziosi	University of Trieste, Italy
Marcela Yepes	Cornell University, USA
Ray Ming	University of Illinois, USA
Lukas Mueller	Boyce Thompson Institute

Agenda ICGN meeting San Diego

1-The second coffee genomics workshop was held January 11, 2009 as part of the XVII PAG meeting. See workshop program at http://www.intl-pag.org/17/17-coffee.html . Abstracts of the oral and poster presentations on coffee at the XVII PAG meeting are included as appendix at the end of this report.

2-During the ICGN meeting held January 12, 2009, the coordinators of the different working groups present gave short updates: Ray Ming for Working group 2, Philippe Lashermes for working group 3, Alex de Kochko for Working group 4, Felipe Rodrigues da Silva and Lukas Mueller for working group 6, and Giorgio Graziosi, as industry representative.

3-Nestor Osorio, Executive Director of the International Coffee Organization (ICO) participated on the ICGN meeting and on discussions with some of the funding agencies to gain an overview of the ICGN efforts towards sequencing the coffee genome.

Summary of the meetings

Second Coffee Genomics Workshop at PAG

More than 50 scientists, including 13 ICGN members, participated in the second coffee genomics workshop held as part of the Plant and Animal Genome Meeting in San Diego January 11, 2009. The co-organizers of the workshop Rod Wing, Philippe Lashermes, and Marcela Yepes appreciated the participation of the invited speakers and their contributions, as well as, the participation of representatives of several agencies including the Specialty Coffee Association of America, the Coffee Quality Institute, the US National Science Foundation (NSF) Plant Genome Program, and representatives of the French National Research Agency (ANR) and Genoscope. Participants to the workshop included Bioversity, and members of the banana sequencing consortium, the cocoa sequencing consortium, the *Solanaceae* community, and Affimetrix, among others.

Program and abstracts of the oral presentations and posters on coffee are included as an appendix at the end of this report. We received very positive feedback from the 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 3rd Coffee Genomics Workshop that will be held January 10, 2010 as part of the XVIII PAG meeting in San Diego, January 9-13, 2010. Please contact one of the organizers if interested on presenting a talk or poster, or with suggestions for new topics for oral presentations or the round table discussion. The coffee genomics workshop is a good opportunity to present advances on coffee genomics research to the

International Plant and Animal Genomics Community and is helping our community explore funding opportunities to sequence the coffee genome.

Round table discussion and participants from funding agencies

The Director of the American Specialty Coffee Association, and the Director of the Coffee Quality Institute kindly accepted our invitation to give short presentations at the end of the workshop to help us justify for the funding agencies the importance of coffee in particular in the US as first coffee consumer country in the world, but also at a global scale as a sustainable crop for many developing countries.

Representatives of ANR-Genoscope attended both the workshop and the ICGN meeting, and also a representative of the NSF Plant Genome Program to discuss interagency funding towards sequencing the coffee genome.

Nestor Osorio, Director of the International Coffee Organization (ICO), kindly accepted our invitation to participate in the ICGN meeting in San Diego. He highlighted the difficulty of viability of the coffee chain in the long term and its vulnerability, and the need to guarantee future sustainability. ICGN will explore with ICO ways to help support our community efforts.

ICGN meeting progress reports from the different working groups present

<u>Report working groups 2 and 4</u>

Last year, a pilot project based on shotgun sequencing using the FLX454 technology was discussed in San Diego, and several ICGN members expressed interest. Research groups of Ray Ming, Alex de Kochko, Georgio Graziosi, and Alan Andrade paid for four runs of 454 GS FLX Titanium for *C. canephora* (using whole genome shot gun WGS sequencing- the cost per run was US $9,000 and the sequencing is on going at the University of Illinois, Urbana). The data and additional markers (SSRs) derived from the 454 *C. canephora* WGS runs are expected shortly.

Contacts are on going with Dominque Crouzillat to utilize the *C. canephora* molecular map developed by Nestlé as an important resource to anchor the coffee physical map. Nestlé has currently 1,150 sequenced based markers mapped to their *C. canephora* map that will be very useful to facilitate genome assembly. Nestlé R&D could communicate to the ICGN members information's (name, location, sequence) related to the mapped markers and genotype data of 93 individual plants constituting the mapping population. Discussions are pending green light from Pierre Broun, new Director of Research & Development at Nestlé. ICGN will ask officially the Indonesian Coffee and Cocoa Research Institute for access to DNA from that population (BP409 X Q121) to be used as international reference genetic map for *C. canephora.* Size of the population can be increased by crossings of the three different parents A x

B, B x C, A x C (each cross has currently a progeny of 93 individuals). ICGN is interested on sharing markers among different groups, and exchange of markers is on going among groups working on different mapping populations.

The Genoscope representative indicated that ideally genetic maps for genome sequence assembling have two sequence-based markers per 1 million base pairs (2 markers/1 Mb). To insure a sufficient density of markers on the whole genome, ICGN will therefore need to map still 750 to 1,000 additional markers on the Indonesian population. All ICGN members involved in working group 2 could contribute to the effort once the new SSR markers/SNP markers are available from the WGS 454 and BAC-ends sequencings. Nestlé, IRD, CIRAD, and University of Illinois have already expressed their willingness to contribute.

Report working group 3

On behalf of working group 3 of the ICGN, CIRAD/IRD submitted a proposal to ANR (France) that was approved in December 2008 for funding. The proposal will allow to BACend sequence at Genoscope the two *Coffea canephora* BAC libraries (*Hin*d III and *Bam*HI) that are being constructed in collaboration with Rod Wing at the Arizona Genomics Institute. CENICAFE and Cornell, on behalf of working group 3 of the ICGN, submitted also a proposal to the InterAmerican Development Bank (FONTAGRO) in 2008. The proposal was approved to construct a BAC library for the diploid maternal parent of *C. arabica* (*C. eugenioides*), BACend sequence the clones, and to fingerprint the *C. canephora* and *C. eugenioides* BAC libraries to construct a coffee physical map as a framework for the sequencing of the coffee genome.

Coordinated interagency funding is currently being search by working group 3 of the ICGN to develop a cost efficient strategy to sequence the coffee genome. The strategy to sequence the coffee genome will involve the adaptation of the novel sequencing technologies (454-Roche and Solexa/Illumina), and will target high coverage *de novo* sequencing to generate reference genomes for both the two diploid parental species *C. canephora* and *C. eugenioides* as well as for the allotetraploid *C. arabica*.

The contacts with ANR (France) are being lead by Philippe Lashermes with a primary phase project already funded to BAC end sequence the *C. canephora* BAC libraries (using Sanger sequencing) and consolidate a reference *C. canephora* genetic map. Development of the *C. canephora* BAC library *Hind*III has been achieved, and BAC end sequencing of 72,000 BAC clones should be done at Genoscope in 2009 with ANR funding. Construction of the 2nd *C. canephora* BAC library is on going in collaboration with Rod Wing at the University of Arizona. Due to technical problems, the restriction enzyme for the 2nd BAC library was switched from *Eco*RI to *Bam*H1. A second proposal (Genoscope, CIRAD, IRD) to ANR will be prepared and submitted in March 2009. If funded,

this proposal will allow to obtain a complete high coverage reference sequence for *C. canephora* using a combination of 454 Titanium (which allows read lengths of 400 bp) /and Solexa-Illumina.

Report Working group 6

Community annotation of the coffee genome should be possible with the expressed sequenced tags (ESTs) collections developed by different groups (>350,000 ESTs). Felipe da Silva highlighted the need to share bioinformatics tools among ICGN members and the importance to have several Bioinformatics Centers around the world. Mirrors can be available for users that do not have large bioinformatics capacity as raw data becomes accessible. Lukas Mueller indicated that the *Solanaceae* community has several Bioinformatics Centers where data is accessible freely and that each database has its own angle depending on the research interests of the different groups. Lukas indicated that through the *Solanaceae* genomics network (http://sgn.cornell.edu/) comparative resources of interest to the coffee community are available.

Another topic discussed is that Genoscope provides only first gene annotation after whole genome sequencing. Therefore, once the project is completed, the genome sequencing data for *C. canephora* should be moved somewhere else. The problem of database longevity apparently has not been solved yet even for the *Solanaceae* community. Therefore, ICGN will have to promote this effort and explore funding through private companies or other resources. Montpellier/Agropolis is considering the creation of a bioinformatics Center to follow up with data analysis after primary annotation.

ICGN membership information

Efficient dissemination of information among ICGN members continues to be an important effort of the network. The Steering Committee discussed this aspect in a separate ICGN meeting held in Campinas, Brazil. In addition, André Charrier informed the ICGN assembly that BioVersity International kindly expressed interest to continue supporting at no-cost the ICGN secretariat.
Please feel free to distribute this ICGN minutes to other scientists or groups that may be interested on becoming ICGN members. A reminder that to become an official ICGN member you should register your membership on-line (http://www.coffeegenome.org/about/members.php).
Please check the ICGN www page and if your name is not yet included please send once more your application to become member (apparently some applications may have been missed un-intentionally). It is very important to our ICGN efforts that people that are interested on being part of the network appear in the mailing list so that they receive information timely. Also, your membership help us demonstrate to funding agencies that the network is truly committed to work at a global scale.

Program Second Coffee Genomics Workshop, San Diego, California

(http://www.intl-pag.org/17/17-coffee.html)

Sunday Afternoon, 11 January, 2009 3:50 pm to 6:00 pm

Coffee Workshop - Pacific Salon 2

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:

3:50 - 4:10
Felipe Rodrigues da Silva , EMBRAPA (Brazil) (felipes@cenargen.embrapa.br)
"Coffest: A Public Resource For Coffea Spp. Est Analysis"
4:10 - 4:30
Dominique Crouzillat , Nestlé (R&D Tours)
(dominique.crouzillat@rdto.nestle.com)
"An Integrated High Throughput SNP And SSR Genotyping Platform Using Hrm Technology For Coffee Mapping"
4:30 - 4:50
Philippe Lashermes, IRD (France) (Philippe.Lashermes@ird.fr)
"Biodiversity and phylogeny in coffee trees (Coffea L.)"
4:50 - 5:10
Carmenza Gongora, CENICAFE (Colombia)
(Carmenza.Gongora@cafedecolombia.com)
"Use of normalized cDNA libraries for enrichment of non-redundant EST sequences and construction of a cDNA array for functional analysis of the coffee transcriptome"
5:10 - 5:30
Carlos Bustamante, Cornell University (cdb28@cornell.edu)
"Genomics and Computational Biology: applications to association mapping and diversity studies"
5:30 - 5:50
Rod Wing, University of Arizona (rwing@Ag.arizona.edu)
"Towards construction of a coffee physical map and sequencing of the coffee genome"

ABSTRACTS 2nd COFFEE GENOMICS WORKSHOP AND COFFEE POSTERS PAG MEETING SAN DIEGO

ORAL PRESENTATIONS

CoffEST: A Public Resource For Coffea spp. EST Analysis

Alex V Barbosa3 , **Raphael MOB Sales**3 , **Alan C Andrade**1 , **Felipe R da Silva**1, 2

1 Embrapa Recursos Genéticos e Biotecnologia, CxP 02.372, Brasília - DF, CEP 70.770-900, Brazil
2 Instituto de Computação, CxP 6176, Campinas – SP, CEP 13.083-970, Brazil
3 Departamento de Ciências da Computação - UnB, CxP 4466, Brasília - DF, CEP 70.910-900, Brazil

EST Genome Projects are a relatively inexpensive way to describe genes. Effectively finding genes of interest, however, is complicated by the overwhelming amount of partial and redundant sequence data generated in such projects. We present here the Web interface to the EST sequence database maintained at Embrapa Recursos Genéticos e Biotecnologia. The database and the interface were originally developed to support the Brazilian Coffee Genome EST project , and latter incorporated *Coffea canephora* EST data contributed by Cornell University (59,721 raw EST sequences, Lin *et al.,* 2005) and Institut de Recherche pour le Développement – IRD (8,782 raw EST sequences, Poncet *et al.*, 2006). It is constantly updated as more data becomes publically available and can be freely accessed at https://alanine.cenargen.embrapa.br/CoffEST. The CoffEST is build by processing the raw chromatogram data and assembling highly similar sequences on UniGenes, as described by Telles and da Silva (2001). UniGene sequences are then compared with all the nucleotide and protein sequences available at GenBank. All data is stored on a PostgreSQL relational database. Searches on this database can be performed based on homology or origin. Homology searches include (i) sequence blast, (ii) Boolean keywords on pre-computed blast results or (iii) browsing the phylogenetic tree of similarities found on GenBank. Origin searches can be as simple as retrieve a sequence by its name (read or UniGene) or the library it came from. More elaborated searches allow one to find UniGenes exclusively, preferentially or differentially expressed on one library or set of libraries.

An Integrated High Throughput SNP And SSR Genotyping Platform Using HRM Technology For Coffee Mapping.

Florent Lefebvre-Pautigny1 , **Michel Rigoreau**1 , **Priyono Priyono**2 , **Dominique Crouzillat**1

1 R&D Nestlé Centre, 101, Av. Gustave Eiffel, Notre Dame d'Oé, B.P. 49716, 37097 Tours CEDEX 2, France
2 Indonesian Coffee and Cocoa Research Institute; Jl. PB Sudirman No. 90 Jember, 68128, Indonesia

Coffea canephora and *arabica* are coffee species with a major economic interest which is also considered as an important raw material for food industry. Therefore this crop species are considered as an orphan crop due to the lack of investments and scientific research from the international community. Among the main priorities the

establishment of a core genetic map for *C. canephora* is certainly of main interest due to high level of diversity in this crop species. The establishment of a high density genetic map on *C. canephora* using mainly SSR and SNP markers was focused on automation of PCR markers using high resolution melting (HRM) technology. The use of HRM for SNP genotyping allows an exhaustive detection of all the SNPs carried on the PCR amplification product facilitating the transposition of the SNPs among different mapping progenies. The detection of SSR polymorphism is more efficient than gel based technologies since the DNA variability is detected immediately after the PCR run without any further step. HRM reliability was estimated using traditional SNP genotyping with TaqMan probe and SSR fluorescent labelling. The results clearly indicate that the HRM method can be used efficiently to detect SNPs and microsatellites DNA polymorphisms in out crossing progenies such as *Coffea canephora*. The resulting coffee genetic map includes more than 700 SNPs and SSRs that are easily transposable on different progenies. This set of markers facilitates the establishment and management of QTLs in a global strategy for coffee marker assisted selection.

Biodiversity And Phylogeny In Coffee Trees (Coffea L.)

François Anthony , **Marie-Christine Combes** , Philippe Lashermes

UMR RPB (CIRAD, IRD, UM2), Centre IRD de Montpellier, BP 64501, F-34394, Montpellier, France

Coffee species belong to the *Rubiaceae* family, one of the largest tropical angiosperm families. All species are perennial woody bushes or trees, which differ greatly in morphology, size and ecological adaptations. Two genera, *Coffea* L. and *Psilanthus* Hook. f., are distinguished based on flowering and flower characteristics. Both genera are naturally found in tropical Africa, *Coffea* occurring also in Madagascar, Grande Comore and the Mascarenes, and *Psilanthus* in South-East Asia, Oceania and North Australia. Particular attention has been paid to the *Coffea* subgenus *Coffea* (103 species), which comprises the majority of species, including those of economic importance, *C. arabica* and *C. canephora*. A molecular phylogenetic analysis of *Coffea* subgenus *Coffea* was undertaken using data of plastid DNA sequences. Sequences of intergenic spacers from a total of 69 *Coffea* species were used for phylogenetic reconstruction using parsimony analyses. The low rate of homoplasy and the low number of characters supporting the main branches confirmed the hypothesis of a rapid and radial mode of speciation in *Coffea* subgenus *Coffea*. Distribution of main clades suggested the Lower Guinea as centre of origin of *Coffea* subgenus *Coffea*. Based on analyses of colonisation of volcanic islands in Indian Ocean and fossil pollen records in a related genus, a recent divergence time of about 150,000-350,000 yr BP was estimated for *Coffea* species.

Use Of Normalized cDNA Libraries For Enrichment Of Non-Redundant EST Sequences And Construction Of A cDNA Array For Functional Analysis Of The Coffee Transcriptome

Carmenza Góngora1 , **Alvaro Gaitán**1 , **Pilar Moncada**1 , **Marco Cristancho**1 , **L. Fernando Rivera**1 , **Carlos E. Orozco**1 , **Robin Buell**3,4 , **Gabriel Cadena**1 , **Herb Aldwinckle**2 , **Marcela Yepes**2

1 Colombian National Coffee Research Center -CENICAFE- Chinchiná, Caldas,

Colombia
2 Department of Plant Pathology, Cornell University, Geneva, NY 14456, USA
3 The Institute for Genomic Research (TIGR), 9712 Medical Center Drive, Rockville Maryland 20850, USA
4 Department of Plant Biology, Michigan State University, East Lansing, MI 48824, USA

A collaborative research project between the Colombian National Coffee Research Center (CENICAFE), Cornell University, and the Institute for Genomic Research (TIGR) co-funded by the Colombian Coffee Growers Federation, the Colombian Ministry of Agriculture, and the US National Science Foundation focused on the construction of three coffee normalized cDNA libraries: one for the allotetraploid *Coffea arabica* (mixed tissue), and two normalized libraries for the diploid species *C. liberica* (leaf and seeds). Normalized libraries enriched for non-redundant expressed sequence tagged (EST) sequences were used in the fabrication of a coffee c-DNA array for functional analysis of the coffee transcriptome. The coffee microarray includes 33,263 cDNAs (19,074 correspond to *C. arabica* and 14,189 to *C. liberica*) and has been used to obtain expression profiling data, using RNA populations from coffee plants from genotypes with differential response to biotic stress (rust and coffee berry borer) for analysis of pathways and novel candidate genes that could be manipulated to enhance resistance. The generation of normalized libraries allowed a broader survey of the *C. arabica* and *C. liberica* transcriptome by enriching for rare mRNAs including transcripts from five different tissues (leaf, roots, seeds, flowers, and callus). Normalization generated a 2X enrichment in unigenes reducing redundancy and increasing array density. Quality control during amplification and sequencing allowed to generate longer cDNAs that were spotted directly on the screening array. The coffee microarray is also being used to identify genes affecting other traits of importance in coffee production.

POSTER ABSTRACTS

SSR Polymorphism In Breeding Populations Of Arabica Coffee With Varying Reactions To Coffee Berry Disease

Chrispine O Omondi1 , **Elijah K Gichuru**1 , **Marie-Christine Combes**2 , **Philippe Lashermes**2

1 Coffee Research Foundation, PO Box 4-00232, Ngenda Road, Ruiru, Central Province, , Kenya
2 Institut de Recherche pour le Développement (IRD), UMR DGPC, Résistance des Plantes, 911, Av. de Agropolis, BP 64501, F-34394, Montpellier, France,

Coffee Berry Disease (CBD) caused by *Colletotrichum kahawae* is the biggest challenge to Arabica coffee production in Africa since 1922 when it was first reported in coffee plantations in Kenya. Breeding for resistance to CBD has gained prominence in many African countries and pre-emptive breeding is also being carried out in some Latin American countries. In Kenya, a breeding programme was launched in 1971 and resistant varieties have been developed. However, a selection method that is efficient and effective, robust enough to identify multiple resistance genes in one plant and versatile to be applied in quality control of a large number of commercial resistant

planting materials is lacking. Search for DNA markers that can fill that void has been initiated at Coffee Research Foundation (CRF) in Kenya and partners. This paper reports polymorphism observed on Microsatellites (SSR) screened for polymorphism between a resistant donor, Rume Sudan, and a susceptible cultivar, SL 28, of fine beverage quality, high yields and good adaptability in Kenya. Seven microsatellites were screened in agarose at CRF while fifteen were screened in poly-acrylamide gel at IRD, France. One and three microsatellites were found to be polymorphic respectively. The polymorphic marker at CRF was analyzed in a breeding population derived from the two parents and the results are presented. The paper further discusses the proposed approach in the search for DNA markers in mapping populations derived from the two parents and their possible application in Marker Assisted Selection (MAS), to complement another mapped gene.

Best Of Both Words: Using Short And Long Reads To Construct UniGenes

Maria Angelica L de Souza1 , **Andre AM Almeida**1 , **Alan C Andrade**2 , **Felipe R da Silva**1, 2 , **Zanoni Dias**1

1 Instituto de Computação, CxP 6176, Campinas – SP, CEP 13.083-970, Brazil
2 Embrapa Recursos Genéticos e Biotecnologia, CxP 02.372, Brasília - DF, CEP 70.770-900, Brazil

New automatic DNA sequencing equipments from 3 companies became available on the last 4 years, with some more coming. Although they are able to generate 200 - 30,000x more total data per run, they all share one unexpected feature: the size of each individual read is shorter. “Traditional” (Sanger) sequencing is able to produce reads of ~1,000 bases, whereas next-gen sequencing produce reads of ~250, 30 or 35 bases (454, Illumina and SOLiD, respectively). Assembling short reads still considered a challenge, especially due to repetitive areas (larger than the read length) on the target sequence.
This work compares the use of 9 different published algorithms to assemble UniGenes using both EST Sanger reads and next-gen sequences on the Brazilian Coffee Genome EST project data: AMOS, Phrap, Celera Assembler, PCAP/CAP3, Velvet, MIRA, VCAKE, SHARCGS and EULER-SR. Internal and external consistency of the resulting assemblies where measured revealing drawbacks of some algorithms. The resulting UniGenes have proven to be extremely useful to find differentially expressed genes.