

International Coffee Genomics Network (ICGN)
Report 10th Coffee Genomics Workshop held at the
XXV Plant and Animal Genome (PAG) Meeting
San Diego, California
January 14-18, 2017

Coffee Genomics Workshop Speakers

1. **Carlos Maldonado, Marcela Yepes, Aleksey Zimin, and Keithanne Mockaitis** Colombian National Coffee Research Center, CENICAFE, Colombia, Cornell University, School of Integrated Plant Sciences, Plant Pathology and Plant-Microbe Biology Section, USA, University of Maryland/ Johns Hopkins University, USA, Indiana University, USA. *Using PACBio Long Reads to Generate High Quality References for the allotetraploid Coffea arabica and its maternal diploid ancestor Coffea eugenoides: characterization of genomic regions containing QTLs for Yield, Plant Height, and Bean Size.*
2. **Lucio Navarro**, Colombian National Coffee Research Center, CENICAFE, Colombia. *Insights from the Genome of the Major Coffee Insect Pest Worldwide: The Coffee Berry Borer.*
3. **Alan Andrade**, Embrapa Café/INOVACAFÉ - UFLA, Lavras-MG, Brazil. *Towards GWAS and Genome Prediction in Coffee: Development and Validation of a 26K SNP Chip for Coffea canephora.*
4. **Luis Felipe Ventorin Ferrão**. Universidad of Sao Paulo (ESALQ/USP), Brazil. *Comparison of Statistical Methods and Reliability of Genomic Prediction in Coffea canephora Populations.*
5. **Kassahun Tesfaye**, Addis Ababa University, Addis Ababa, Ethiopia. *Coffee Forest Biodiversity and Implications for Multi-Site in situ Conservation Approach in the Afromontane Rainforests of Ethiopia.*
6. **Allen Van Deynze**. University of California Davis, USA. *Update on the Sequencing of the Coffea arabica Variety, Geisha.*
7. **Marco Cristancho**, Colombian Center for Bioinformatics and Computational Biology (BIOS), Manizales, Colombia: *The Colombian Center for Bioinformatics and Computational Biology.*

See abstracts of all presentations included at the end of this report.

Coffee Genomics Workshop at PAG

The Plant and Animal Genome (PAG) meeting celebrated this year a major milestone : its 25th Meeting! PAG continues to be the world largest international scientific conference reporting on animal and plant genomics advances, with >3,000 participants from >65 countries around the world. For those interested in participating in future PAG meetings see <http://www.intlpag.org>. The XXVI Plant & Animal Genome Conference will be held in San Diego, January 13-17, 2018.

A record number of scientists participated this year in our 10th ICGN coffee genomics workshop held as part of the PAG Meeting in San Diego at the Pacific Salon 3, on January 15, 2017. We marked this important milestone for ICGN by featuring a short historic summary of prior coffee genomics workshops at PAG highlighting the major achievements of our coffee genomics community over the past decade. The co-organizers of the workshop, Marcela Yepes (Cornell University, my11@cornell.edu), and Philippe Lashermes (IRD-CIRAD, France, philippe.lashermes@ird.fr), thank the speakers and workshop participants for their contributions. In addition, special thanks to Mr. Jay Ruskey, owner of Good Land Organics (www.goodlandorganics.com) who brought as display several coffee plants, *Coffea arabica* var. Geisha, and kindly featured coffee tasting of this coffee variety for our workshop participants.

Over the past decade our coffee genomics community has focused efforts on bringing coffee to the forefront of plant genomics research. Abstracts of our 10th coffee genomics workshop presentations are included as an appendix at the end of this report. Please mark your calendars for our 11th Coffee Genomics Workshop that will be held Sunday January 14, 2018 during the XXVI PAG meeting in San Diego, January 13-17, 2018. 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 collaboration with the International Coffee Organization (ICO) and participation in the First World Coffee Producers Forum organized by the Colombian Coffee Growers Federation

ICGN sent condolences to the ICO for the sudden passing of its Executive Director Dr. Robério Oliveira Silva on December 30, 2016. ICGN also sent condolences to the Brazilian colleagues and International coffee community on this loss. ICGN congratulates the naming of the new ICO Director José Dauster Sette and supports ICO's efforts to promote global coffee sustainability with emphasis in economic, social and environmental sustainability. We also congratulate the Colombian National Coffee Growers Federation (FNC) in its celebration this year of its 90 years by organizing the first World Coffee Producers Forum in Medellín, Colombia, from July 10-12, 2017. The forum brought together 1,300 participants from 44 countries to analyze major challenges to the coffee value chain including: coffee farmers' economic sustainability, productivity, price volatility, the next generation of coffee growers, the consequences of climate change, how to increase quality and traceability for consumers and the increasing demand of 50 million bags in the next 10 to 15 years. The forum addressed principles of global co-responsibility and cooperation needed for future sustainability of the coffee chain.

ICGN shares the priorities that were discussed at the first Coffee Producers Forum and will continue to network the world coffee genomics scientific community to help prioritize and advance topics of major relevance for future sustainability of the coffee chain.



The Forum aimed at building, with the help of representatives of the different links (coffee producers, exporters, importers, roasters, traders, etc.), an accurate diagnosis of the main challenges of the coffee value chain to achieve future sustainability for the benefit of everybody, starting by producers and involving consumers. The forum addressed the main challenges and problems of coffee farming to identify strategies for a better distribution of value and responsibility among all actors, in search of sustainability of the global chain from seed to cup. Because of its significance, the forum was without doubt one of the most important events of the global coffee industry in recent years, and was attended by US former President Bill Clinton, Colombian President Juan Manuel Santos, Honduras President Juan Orlando Hernández, Costa Rican President Luis Guillermo Solís, and ICO's Executive Director José Dauster Sette, all pictured above. The Forum was organized and hosted by the Colombian National Coffee Growers Federation (FNC) (see FNC's CEO Roberto Vélez pictured above right from the iconic Juan Valdez who represents since 1958 the Colombian coffee farmers).

ICGN survey and networking efforts

ICGN is conducting a survey 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 interested on collaborating in these efforts are asked to help us contribute by completing and submitting the survey available at our www site (<http://www.coffeegenome.org>). Survey results will be discussed at the next ICGN meeting held in conjunction with the 2018 PAG meeting in San Diego.

As the first *de novo* coffee genome references and assemblies become publicly available (see update on the status of on-going projects included this report), 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 genomic tools. Advanced resources in coffee genomics will help address challenging issues for our

community such as climate change adaptation and sustainability that could be accelerated with transforming genomic tools 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 and development pressures in its native habitat. In addition with support from the International Coffee Organization (ICO), ICO member countries have been contacted to discuss possible interest on 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 continuous support by previous ICO Executive Directors Dr. Nestor Osorio and Dr. Robério Oliveira Silva, who invited ICGN to participate as an observer in the ICO Council meetings. We look forward to working closely with ICO officials under the leadership of the new Executive Director José Dauster Sette on the preparation and submission of a first ICGN/ICO proposal, and to explore potential sources of finance for such joint initiatives. Support from ICO and the private sector will be key for ICGN to secure future funding for diversity conservation efforts in *Coffea* with a broader funding base, and to promote coffee genomics research for coffee improvement targeting priority traits for the coffee industry. Capacity building in developing countries to participate in coffee genomic research is being supported through ICGN networking efforts and our yearly workshops at PAG.

ICGN is particularly grateful to Bioversity International and IRD/CIRAD for supporting our networking efforts by hosting at no cost to our community the ICGN website since its inception in 2005. Bioversity and IRD/CIRAD reiterated in 2017, their interest to help us give continuity to the ICGN www site for the benefit of the international coffee scientific community.

Coffee importance

Coffee is the world's most valuable agricultural export commodity, second only to oil in international trade (International Coffee Organization, <http://www.ico.org>), with 17 billion pounds of green coffee produced around the world each year on over 25 million acres. Global coffee consumption continues to increase at an annual rate of 2.5-3%. About 125 million people in more than 60 developing countries depend on coffee for their food security and livelihoods. In some countries rural employment in coffee

production accounts for 80% of foreign trade earnings, and nearly 75% of global coffee production comes from small farms of less than 5 acres. The retail value of the coffee industry is estimated to be 90 billion US dollars per year and yet, remarkably, coffee remains an understudied crop, receiving very little international funding for genetics and genomics research.

For most coffee growing countries, coffee production is a powerful job creator and a major economic driver. In Africa, eight out of 25 coffee producing countries are among the ten poorest countries in the world, and 53% of the rural population on those 25 coffee producing countries, are involved in coffee growing. Of major concern to the global coffee chain is that coffee production in Latin America, Africa, and around the globe continues to decline due to biotic and abiotic stresses, with increasing temperature, excessive rain or draught affecting directly the incidence of pests and diseases. ICGN targets the development of advanced genomic tools for coffee as a new approach to revitalize coffee production on a global scale and tackle the global challenges facing production for the coffee sector in the context of climate change. Attaining efficient production to ensure sustainability of the coffee sector is a major goal for the coffee community worldwide. ICGN's long-term goal is to address key issues such as adaptation of the crop to climate change and help in the transformation of coffee production at a global scale with innovative genomic tools that accelerate linkage of genotypic and phenotypic diversity in coffee.

Update status of the *Coffea canephora* genome sequencing

With funding from the Agence Nationale de la Recherche (ANR; Genoplante ANR-08-GENM-022-001), 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 included 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). 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 our 4th ICGN Coffee Genomics Workshop at PAG in San Diego in 2011. In 2013, France Denoeud from Genoscope presented at our 6th ICGN Coffee Genomics Workshop an update on the first genome assembly, and in 2014 Alexis Dereeper presented during our 7th ICGN Coffee Genomics Workshop the **Coffee Genome Hub**, an integrative genome information system accessible through the South Green Bioinformatics Platform, developed to provide centralized access to all the coffee scientific community of the full *C. canephora* genome sequence, as well as genomics, genetics, mapping, and breeding data and analysis tools to facilitate basic, translational and applied research in coffee (Dereeper, et al. 2014. Nucleic Acids Research. 43: D1028-D1035, access free manuscript at <http://nar.oxfordjournals.org/content/43/D1/D1028.full.pdf+html>) . **The**

manuscript for the sequencing of the *C. canephora* genome was published in 2014 (Denoed *et al.* 2014. Science 345: 1181-1184); access manuscript at <http://www.sciencemag.org/content/345/6201/1181.full>, and the genome assembly can be freely accessed at: <http://coffee-genome.org>.

Update status of the *Coffea eugenioides* genome sequencing

This project was funded by the Inter American Development Bank, (FONTAGRO/SECCI), with co-funding from the Colombian National Coffee Growers Federation (FNC) and its National Coffee Research Center, CENICAFE. Additional funding for the project was secured in 2016 from the US National Science Foundation (NSF). Genome sequencing for *Coffea eugenioides* was started towards the end of 2012. The project has been developed collaboratively by CENICAFE and Cornell University. Funding for this project was secured jointly through a proposal prepared and submitted by Cornell University and FNC/CENICAFE.

We mimicked 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 collaborated with Roche to construct and sequence 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 670 Mb). We collaborated with Roche to construct and sequence twelve 20 Kb long insert libraries (3.1X coverage) using paired end sequencing and Roche 454 FLX Titanium. The first genome assembly for *C. eugenioides* using the 454 data described above and Newbler v.3.0 assembler was completed in collaboration with Roche in 2014, and was presented by Marco Cristancho from CENICAFE at our 8th ICGN Coffee Genomics Workshop and at the Roche Workshop during the PAG meeting in San Diego in January, 2015. Below is a summary table comparing the initial raw assembly for *C. eugenioides* with the initial raw assembly for *C. canephora* (statistics from Table S2 of manuscript Denoed *et al.* 2014, and from data presented by Genoscope at PAG 2013):

Initial Raw Assembly Summary	<i>C. eugenioides</i> Heterozygous	<i>C. canephora</i> Homozygous Double haploid
No. of contigs (>100 bp)	364,530	211,157
No. of contigs (>500 bp) raw assembly	146,520	96,182
No. of scaffolds	30,263	13,345
Average scaffold size	15,897 bp	42,606 bp
Genome size assembled	508 Mb	569 Mb
Estimated genome size	630-660 Mb	710 Mb
% genome assembled	77-80.6%	80.14%
N50	209,891 bp (502 scaffolds)	1,260,636 bp (108 scaffolds)
Largest scaffold	4.0 Mb	9.0 Mb
Assembler Newbler	v.3	v.2.3
Q40	97.85%	
Inferred read error	1.41%	
Percent repeats	7%	

In addition, we have constructed and sequenced a *C. eugenoides* library using PACBio P6/C4 chemistry (read N50 16 Kb) to 59X genome coverage, and generated also Illumina Moleculo (synthetic 10 Kb fragments) to help us connect and reduce the overall number of contigs and scaffolds in the *C. eugenoides* assembly, as well as to increase the overall percent of genome assembled. Error curation of the *C. eugenoides* PACBio genome assembly using high coverage PCR free Illumina (125.7X 250 bp paired end reads) prior to annotation was presented at our 2017 coffee genomics workshop by our NSF co-Principal investigator Aleksey Zimin, from University of Maryland/ Johns Hopkins University. Anchoring of the contigs to chromosomes using molecular genetic mapping data and Dovetail Hi-C, and annotation are on going. The reference genomes of the diploid species *C. canephora* and *C. eugenoides* (parental diploid ancestors of the allotetraploid species *Coffea arabica*) will be public resources and serve as frames for assembly of *C. arabica*, the major cultivated coffee species worldwide.

Update *Coffea arabica* sequencing

With funding from the Inter American Development Bank (IDB/FONTAGRO) and more recently (2016) from the US National Science Foundation (NSF), Cornell University and FNC/CENICAFE sequenced also the allotetraploid *Coffea arabica* genome generating a very high quality deep coverage PACBio only assembly of *C. arabica* (69-73X genome coverage); N50 read length 12-15 Kb (Average 12 Kb; longest read 65 Kb). The allotetraploid assembly was error corrected using high coverage PCR free Illumina (143X, 250 bp paired end reads) prior to annotation. Validation of the allotetraploid assembly using the genome assemblies of its two diploid ancestral species is on going. We have also generated Illumina Moleculo (synthetic 10 Kb fragments) data for *C. arabica* that has been used to validate the PACBio only assembly. Transcriptome assemblies and the recently published *Coffea arabica* high density molecular genetic map, developed by CENICAFE in collaboration with Cornell University (Moncada *et al.* 2016 Tree Genetics and Genomes 12: 5 DOI 10.1007/s11295-015-0927-1), as a well as a physical map developed by CENICAFE in collaboration with Dr. Rod Wing at University of Arizona, and Dovetail Hi-C data are being used to validate the high quality *de novo* genome assembly and generate chromosome level scaffolds. On going work to characterize QTLs of interest was presented at our 2017 coffee genomics workshop by Carlos Maldonado from CENICAFE. This very high-quality *C. arabica* reference genome assembly should dramatically improve our current understanding of coffee genetics and genomics providing direct applications to breeders for climate change adaptation. Integration of genomic studies of equivalent quality among the allotetraploid *C. arabica* and its diploid progenitors will maximize scientific insights into the complex biology of polyploids.

Updates on the status of the *C. arabica* sequencing projects funded by private companies, Nestlé and IllyCaffé/Lavazza, were presented at our 8th ICGN Coffee Genomics Workshop at PAG in 2015 (see report in our ICGN www site) and for a more recent update see abstracts of the ASIC meeting in Yunnan, China in 2016 at <http://www.asic2016china.org/>. Update on the effort funded by Suntory for the genome

sequencing and assembly of the *C. arabica* variety Geisha was presented at our 2017 workshop (see abstract below).

Perspectives High quality coffee genome and transcriptome assemblies for *C. arabica* and its two diploid ancestral species are important resources to accelerate future coffee improvement strategies. They are helping us in the identification of genes involved in important agricultural traits in coffee, as we build up crucial information on the structural variation between *Coffea* wild species and cultivated accessions. They are important template for positional cloning of agriculturally important genes, for re-sequencing and deep diversity analysis in the *Coffea* gene pool, as well as an important support to develop genomic tools for whole-genome expression analysis. They will also provide a solid foundation to study the evolution of euasterids and accompanying genomic changes.

To ensure full benefit from the generated coffee genomic sequences and resources by the coffee sector, ICGN continues to explore additional funding from International Funding Agencies to support our coffee genomics scientific community efforts.

Abstracts of Oral Presentations 10th ICGN Coffee Genomics Workshop at PAG co-organizers

[Marcela Yepes](mailto:my11@cornell.edu), Cornell University (my11@cornell.edu)

[Philippe Lashermes](mailto:philippe.lashermes@ird.fr), L'Institut de Recherche pour le Développement
(IRD), France (philippe.lashermes@ird.fr)

(Program and abstracts and pdfs of some of the presentation are also posted at:
<https://pag.confex.com/pag/xxv/meetingapp.cgi/Session/4188>)

Using PACBio Long Reads to Generate High Quality References for the allotetraploid Coffea arabica and its maternal diploid ancestor Coffea eugenioides: characterization of genomic regions containing QTLs for Yield, Plant Height, and Bean Size.

Carlos Ernesto Maldonado¹, Marcela Yepes², Aleksey Zimin^{3,4}, Keithanne Mockaitis⁵, Alvaro Gaitán¹, Pilar Moncada¹, Carrie Ganote⁵, Beatriz Elena Padilla⁶, Narmer Fernando Galeano⁷, Carmenza E. Góngora¹, Claudia Flórez¹, Ricardo Acuña¹, Andrés Mauricio Villegas¹, Huver Posada⁸, Rod A. Wing⁶, James A Yorke³ and Herb Aldwinckle².

¹Centro Nacional de Investigaciones de Café, CENICAFE, Chinchiná, Colombia, ²Cornell University/ School of Integrative Plant Sciences/ Plant Pathology and Plant-Microbe Biology Section, Geneva, NY, ³University of Maryland, College Park, MD, ⁴Johns Hopkins University, Department of Computer Science, Baltimore, MD, ⁵National Center for Genome Analysis Support, Indiana University, Bloomington, IN, ⁶Arizona Genomics Institute, University of Arizona, Tucson, AZ, ⁷Universidad Católica de Manizales, Manizales, Colombia, ⁸Federación

Nacional de Cafeteros de Colombia (FNC)/ Centro Nacional de Investigaciones de Café (CENICAFE), Chinchiná, Caldas, Colombia.

We are celebrating this year our 10th ICGN Coffee Genomics Workshop at PAG. Over the past decade our coffee genomics community has focused efforts on bringing coffee to the forefront of plant genomics research.

The first coffee genome assembly published (Denoed *et al.* 2014) was for the diploid cultivated species *Coffea canephora*. In parallel, Cornell University and FNC/CENICAFE, submitted a proposal to IDB/FONTAGRO to sequence the genomes of the most widely cultivated coffee species, the allotetraploid *Coffea arabica*, and its diploid maternal ancestor *C. eugenoides*. We received funding in 2016 from NSF to strengthen this effort and generate state of the art high quality reference genomes for these *Coffea* species to help accelerate linkage of structural and functional diversity in coffee for climate change adaptation. We used high coverage PACBio long reads for *de novo* genome assembly, and PCR free Illumina paired-end sequencing data for error correction using the MaSuRCA genome assembler (Zimin *et al.* 2017) prior to genome annotation and high through put diversity variant calling. We built *de novo* transcriptome assemblies for *C. arabica* using RNA reads of a variety of functional genomics experiments, and used these to progressively validate completeness and quality of our genome assemblies.

In addition, progress in the characterization of genomic regions containing QTLs for yield, plant height, and bean size in *C. arabica* through the integration of the linkage groups harboring the QTLs (Moncada *et al.* 2016 Tree Genetics and Genomes 12: 5 DOI 10.1007/s11295-015-0927-1), the physical map of the species (contracted by CENICAFE to Rod Wing at U. of Arizona), genomic sequences obtained by whole genome and targeted sequencing (BAC by BAC), and full length transcriptome (Iso-Seq method, PACBio) from the progenitors of the population used to detect the QTLs will be presented. This analysis will provide candidate genes and genomic features related to important agronomic traits in the allotetraploid *C. arabica*, useful for functional genomics and to develop tools for marker assisted selection.

The PACBio coffee genome assemblies were done in collaboration with Pacific BioSciences and the Colombian Center for Bioinformatics and Computational Biology (Bios).

This abstract had an extended time (40 min) and was presented by co-authors M. Yepes (project introduction PACBio assemblies), A. Zimin (hybrid genome assemblies error correction), K. Mockaitis (transcriptome assemblies), and C. Maldonado (characterization genomic regions containing QTLs for traits of interest).

This project is co-funded by the US National Science Foundation, the InterAmerican Development Bank, and the Federación Nacional de Cafeteros de Colombia through its National Coffee Research Center, CENICAFE.

Insights from the Genome of the Major Coffee Insect Pest Worldwide: The Coffee Berry Borer.

Lucio Navarro¹, Flor Edith Acevedo¹, Jonathan Nuñez¹, Erick Hernandez², Rita Daniela Fernandez-Medina³, Claudia Carareto⁴, Pablo Benavides¹ and Stuart Jeffrey⁵.

¹Centro Nacional de Investigaciones de Cafe, CENICAFE, Chinchiná, Colombia, ²Sao Paulo State University, Sao Jose do Rio Preto, Brazil, ³Escola Nacional de Saúde Pública, Fundação Oswaldo Cruz, Rio de Janeiro, Brazil, ⁴UNESP - Universidade Estadual Paulista, São José do Rio Preto, SP, Brazil, ⁵Purdue University, West Lafayette, IN, USA

The Coffee Berry Borer (CBB, *Hypothenemus hampei*) brings major challenges for insect control due to its particular biology and genetics. Most of its life cycle occurs inside the coffee bean where an extreme inbreeding drives the mating behavior between the diploid females and their parahaploid male siblings. These CBB biological features make hard to implement effective control methods in field conditions. The availability of CBB whole genome assemblies open new opportunities to better understand the biology of the insect and its interaction with coffee plants. A hybrid *de novo* CBB genome assembly (~160 Mb) using FLX-454 and Illumina reads from both female and male individuals was presented. Compared with our initial FLX-454-based assembly and other published CBB genome assembly, the new hybrid assembly has improved sequence contiguity. Transcriptomics data obtained from RNA-seq supported around 21,000 predicted genes in this assembly, which account for over 95% of genome completeness. We annotated different gene families of interest, including odorant receptors and odorant-binding proteins as well as G protein coupled receptors (GPCRs) as a prerequisite for exploring new methods of insect behavioral control or selection of safer insecticides. A reduction in genes related to olfactory functions was found comparable with other curculionid beetles. Genome sequence analyses revealed also a low content of repetitive DNA compared with other insect genomes. Only ~8% of the CBB genome consist of transposon elements and ~1% of tandem repeats. This low content of repetitive DNA sequences may represent an evolutionary adaptation to the extreme inbreeding in the CBB. Female and male-specific genome assemblies showed structural differences. This information along with the identification of several genes involved in sex determination mechanisms are essential to elucidate the sex-determination process in the insect.

Towards GWAS and Genome Prediction in Coffee: Development and Validation of a 26K SNP Chip for *Coffea canephora*.

Alan C. Andrade¹, Orzenil B. da Silva Junior², Fernanda de A. Carneiro³, Pierre Marraccini^{4,5} and Dario Grattapaglia².

¹Embrapa Café/INOVACAFÉ - UFLA, Lavras-MG, Brazil, ²Embrapa Recursos Genéticos e Biotecnologia, Brasília-DF, Brazil, ³Graduate Program on Plant Biotechnology - UFLA, Lavras-MG, Brazil, ⁴Embrapa Recursos Genéticos e Biotecnologia, Brasília-DF, Brazil, ⁵CIRAD UMR AGAP, Montpellier, Brazil.

Genome-wide SNP genotyping platforms aiming at high-throughput and high-precision

genotyping constitute an essential tool to advance breeding by genomic prediction and gene discovery by GWAS. Recent advances in coffee genomics with the sequencing of the *Coffea canephora* reference genome, has provided the coffee scientific community the necessary resource to develop a SNPs toolbox for genome-wide genotyping. *C. canephora*, an allogamous diploid species, and one of the parents of the allotetraploid *C. arabica*, has been an important source of genetic variability for breeding programs of both cultivated species. Highly heterozygous genomes such as *C. canephora* require a much higher sequence depth to reach acceptable marker call rates and genotype accuracy, when using sequence-based genotyping methods such that their cost effectiveness may not be realized. Here we describe the development and validation of a 26K Axiom SNP array (Affymetrix) whose genome-wide distributed SNP content was discovered from pooled whole-genome resequencing of *C. canephora* accessions covering most of its known genetic diversity. Besides facilitating low cost, high marker density, polymorphism and speed of data generation, the platform displays high genotype call accuracy and reproducibility. Genotyping validation resulted in 23,585 SNPs (92.6%) successfully converted out of the 25,456 SNPs on the array and 19,586 of them (83%) were deemed “high-resolution polymorphic in a set of 800 individuals of a breeding population. This large validated SNP collection provides a powerful tool for molecular breeding and population genetics investigation within coffee species. Some preliminary results of a GWAS using this genotyping platform will be presented.

Comparison of Statistical Methods and Reliability of Genomic Prediction in *Coffea canephora* Populations

Luis Felipe V. Ferrão¹, Romario G. Ferrão², María A. G. Ferrão³, Aymbire Fonseca³, Matthew Stephens⁴, and Antonio Augusto Franco Garcia¹

¹University of São Paulo (ESALQ/USP), PIRACICABA, Brazil

²Instituto Capixaba de Pesquisa, Assistência Técnica e Extensão Rural, Vitoria, Brazil

³Instituto Capixaba de Pesquisa, Assistência Técnica e Extensão Rural / Embrapa Café/ Vitória-ES, Brazil

⁴University of Chicago, Chicago, IL, USA

Simulation and empirical results have shown that genomic predictions present sufficient accuracy to help increase success in breeding programs. Although many crops have benefited from this novel approach, studies in the *Coffea* genus are still in their infancy. Until now, there have been no studies of how predictive models work across populations and environments or, even, their performance for different complex traits. Considering that predictive models are based on biological and statistical assumptions, it is expected that their performance varies depending on the true underlying genetic architecture of the phenotype. We used real data from two experimental populations of *Coffea canephora*, evaluated in two environments (sites) and SNPs identified by Genotyping-by-Sequencing (GBS) to investigate the genotype-phenotype relationship. We considered Bayesian models, with different prior distributions for the marker effects, and regularized linear regression models. We assessed predictive abilities using a Replicated Training-Testing evaluation, with 30 repetitions, and different metrics to compare the model performances. In addition, we investigated SNP effects to learn about underlying biology related to

genomic regions affecting the phenotype and their interactions. For the three traits evaluated, there were minimal differences in predictive accuracy among models. A slight advantage of Bayesian methods was observed, although more computation was required. Predictions within-population, on average, were more accurate than between populations. Biological insights revealed genetic variants with specific signals within populations and environments. Consequently, these results have great potential to reshape traditional breeding programs, including genomic predictions for improved breeding strategies.

Coffee Forest Biodiversity and Implications for Multi-Site *in situ* Conservation Approach in the Afromontane Rainforests of Ethiopia

Kassahun Tesfaye¹, and Feyera Sembeta²

¹Addis Ababa University, Addis Ababa, Ethiopia

²College of Development Studies Addis Ababa University, Addis Ababa, Ethiopia

Arabica coffee (*Coffea arabica*) originates in the montane forests of southwest and southeast Ethiopia. Recently these forests have come under continuous threat due to anthropogenic factors. A study was conducted to assess plant species and coffee genetic diversity in five forest fragments. A total of 651 species that belong to 118 families were recorded from five forest fragments. Among the species recorded, about 5% are endemic plants. Of the total species recorded about 50% of the species occur in only one of the forests indicating the uniqueness of the forests. Diversity of Arabica coffee was assessed using ISSR and AFLP markers system. These analyses showed a complex pattern of genotype distribution; whereby individuals from some regions spread all over the trees generated whereas others form their own groups. Moreover, higher diversity within populations of *C. arabica* was also evidenced with unique genotypes from each forest. These results from both floristic and genetic diversity suggest the need for multi-site *in situ* conservation approach to capturing the diversity and uniqueness found in different wild coffee regions. In addition to the conservation effort, advanced genomic tools should be applied for in-depth diversity study and trait discovery for conservation and sustainable use of Arabica coffee genetic resources in montane rainforests of Ethiopia.

Update on the Sequencing of the *Coffea arabica* Variety, Geisha.

Allen Van Deynze¹, Amanda M. Hulse-Kemp², Michael C. Schatz³, Jason Chin⁴, Jay Ruskey⁵, Dario Cantu¹ and Juan F. Medrano¹

¹University of California, Davis, CA, ²University of California, Raleigh, NC, ³Johns Hopkins University, Baltimore, MD, ⁴Pacific Biosciences, Menlo Park, CA, ⁵Good Land Organics, Goleta, CA

Coffee traditionally is grown worldwide at equatorial latitudes below 25° under very specific growing conditions of acid soils, warm temperatures and high humidity. The environment has a direct effect on the quality and final taste of the berry. The variety

Geisha originates from the mountains of the western Ethiopian provinces of Maji and Goldija, near the town of Geisha, and is a selection known for its unique aromatic qualities. Over the last 6 years, this variety has been successfully grown near Santa Barbara, California, 19 ° latitude north of any other plantation. We have sampled and sequenced DNA and transcriptomes from this variety. RNA samples from different tissues and developmental stages were collected and sequenced to enhance gene model prediction in combination with *ab initio* methods. Functional annotations focused on pathways relevant to coffee quality and adaptation to biotic and abiotic stresses. Resequencing of a panel of 15 Geisha accessions will provide a first glimpse on the genetic variation within this variety and an additional 10 varieties. An understanding of diversity within and among varieties at the whole genome level will be presented. Annotations, structural variants and polymorphisms in candidate genes and pathways associated with coffee quality are being investigated to understand the flavor profiles of Geisha coffee.

BIOS: *The Colombian Center for Bioinformatics and Computational Biology*

Marco Cristancho, Colombian Center for Bioinformatics and Computational Biology (BIOS), Manizales, Colombia.

The Colombia National Center for Bioinformatics and Computational Biology-BIOS is a leading institute in Latin America in the study of the rich and unique biodiversity found in countries of the region. South and Central America are the centre of origin, diversity and domestication of several economically important crop species, including maize, tomato, and potato. Most of the biodiversity in Latin American countries is under-studied and very few endemic plant species in the region have been sequenced at any level. We work closely with Research Institutes in Colombia and neighboring countries, promoting projects for the development of plant genomics and bioinformatics research and resources to forward scientific and economic development for the region.

BIOS is driving bioinformatics research in Latin America, incorporating high quality standards for data acquisition, sequence analysis, data storage, and facilitates data access through visualization platforms. The Centre is already gathering sequence data from several plant species endemic to the Andes, Amazon and the Orinoquia regions of Colombia, while working closely with the health, food, and cosmetic industry in the development of novel products from those plants. Our genomics and bioinformatics studies are coordinated with major Research Initiatives such as Colombia BIO, a national endeavour to study and use the vast biodiversity of the country in a sustainable way.

We are also supporting efforts to increase agricultural productivity for crops of economic importance in the region. By searching for genes related with increased productivity, disease resistance, dry tolerance and other characteristics that can make plants resilient to climate change. We have been carrying these studies with leading Research Institutes, including the International Center for Tropical Agricultural – CIAT, and the Colombian Center for Sugarcane Research– CENICAÑA.

Given the great economic importance of coffee production to Colombia and the region, we are also collaborating closely with Cornell University, the Federación Nacional de Cafeteros de Colombia (FNC) and its National Coffee Research Center (CENICAFE) in the *de novo* coffee genome assemblies for the species *Coffea arabica* and *C. eugenoides* using Pacific Biosciences as well as other sequencing technologies.

Pictures of our ICGN 10th Coffee Genomics Workshop Speakers and Participants at XXV PAG 2017



Aleksey Zimin, University of Maryland/ Johns Hopkins, **Herb Aldwinckle**, Cornell University, **Keithanne Mockaitis**, Indiana University, and **Marcela Yepes**, Cornell University.



Marco Cristancho, Colombian Center for Bioinformatics and Computational Biology, BIOS, **Marcela Yepes**, Cornell University, **Keithanne Mockaitis**, Indiana University, **Kasshun Tesfaye**, Addis Ababa University, Ethiopia, **Sherie Sanders**, **Carrie Ganote**, and **Thomas G. Doak**, NSF funded National Center for Genome Analysis Support (NCGAS) at Indiana University.



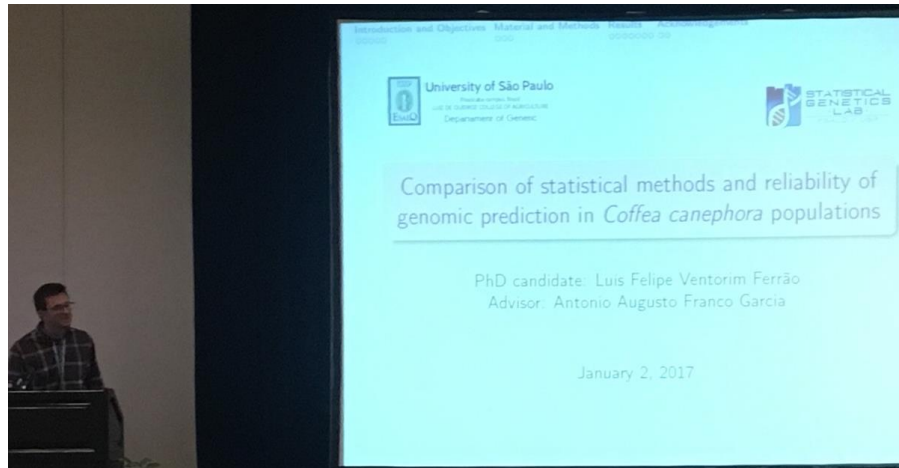
Carlos Maldonado, Colombian National Coffee Research Center, FNC/CENICAFE, Colombia



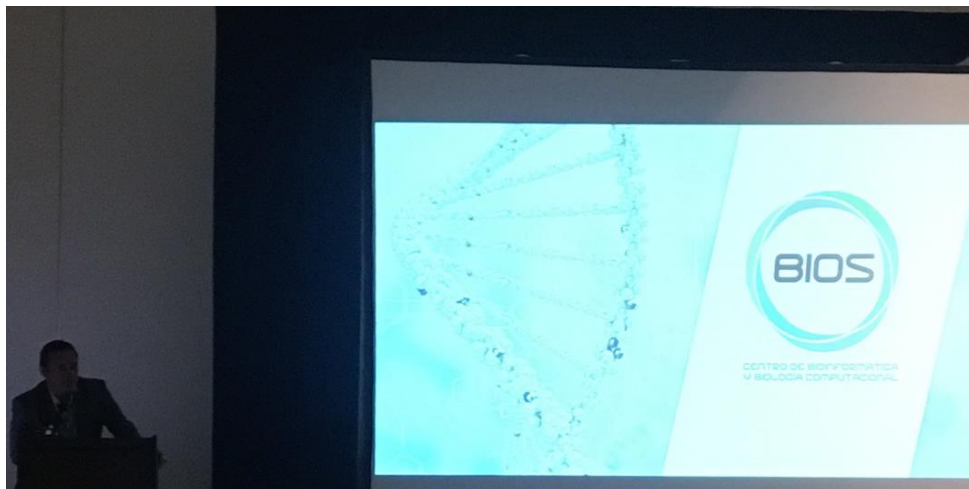
Lucio Navarro, Colombian National Coffee Research Center, FNC/CENICAFE, Colombia



Alan Andrade, Embrapa Café/INOVACAFÉ - UFPA, Lavras-MG, Brazil.



Luis Felipe Ventrin Ferrão. Universidad of Sao Paulo (ESALQ/USP), Brazil.



Marco Cristancho, Colombian Center for Bioinformatics and Computational Biology (BIOS).



Kassahun Tesfaye, Addis Ababa University, Addis Ababa, Ethiopia.



Aleksey Zimin, University of Maryland/ Johns Hopkins University, Dep. of Computer Sciences, USA.



Keithanne Mockaitis, NSF-funded National Center for Genome Analysis Support (NCGAS), and Biology Department, Indiana University.



Allen Van Deynze, University of California, Davis.



Allen Van Deynze, **Amana Hulse-Kemp**, and **Juan Fernando Medrano**, University of California, Davis, with **Jay Ruskey**, owner of Good Land Organics (www.goodlandorganics.com), who brought for display several potted coffee plants (*Coffea arabica* variety Geisha) and kindly provided coffee tasting of this variety for our coffee genomics workshop participants.



Keithanne Mockaitis, Indiana University, **Marcela Yepes**, Cornell University, **Fernando Medrano**, University of California, Davis.



Kassahun Tesfaye, Addis Ababa University, Ethiopia; **Marcela Yepes**, Cornell University, **Juan Fernando Medrano**, University of California, Davis.



We thank Pacific BioSciences for hosting our scientists in their social event at PAG.



Ray Ming, University of Illinois, Urbana, **Herb Aldwinckle**, Cornell University and **Kassahun Tesfaye**, Addis Ababa University during the Pacific BioSciences hosted social.



Luis Felipe Ferrão and **Juliana Benevenuto**, Universidad of Sao Paulo (ESALQ/USP), Brazil.

Recent and Upcoming Meetings of interest to the ICGN community

- 1st World Coffee Producers Forum, organized by the Colombian National Coffee Growers Federation as part of the celebration of its 90th years. Medellín, Colombia, July 10-12, 2017. <https://www.worldcoffeeproducersforum.com/> This forum brought together 1,300 participants from 44 countries to identify strategies for a better distribution of value and responsibility among all actors of the coffee chain including the coffee global industry, in search of sustainability of the global chain from seed to cup.
- 14th Solanaceae Genomics Network SOL Meeting, Valencia, Spain, 2017, September 3-6, 2017, solcuc2017.org / Abstracts of 2016 13th Sol meeting are available at <http://solgenomics2016.ucdavis.edu/>
- 11th ICGN Coffee Genomics Workshop at XXVI Plant and Animal Genome (PAG) Meeting, San Diego, California, January 13-17, 2018, <http://www.intlpag.org/>
- 28th ASIC International Conference on Coffee Science, Portland, Oregon, USA, 2018, September 16-20, 2018, <http://www.asic-cafe.org>