Huver Posada[8], Rod A. Wing[6], James A Yorke[3] and Herb Aldwinckle[9], (1)Centro Nacional de Investigaciones de Café, CENICAFE, Chinchiná, Colombia, (2)Cornell University/ School of Integrative Plant Sciences/ Plant Pathology and Plant Microbe Biology Section, Geneva, NY, (3)University of Maryland, College Park, MD, (4)Johns Hopkins University, Department of Computer Science, Baltimore, MD, (5)Indiana University, Bloomington, IN, (6)Arizona Genomics Institute, University of Arizona, Tucson, AZ, (7)Universidad Católica de Manizales, Manizales, Colombia, (8)Federación Nacional de Cafeteros de Colombia (FNC)/ Centro Nacional de Investigaciones de Café (CENICAFE), Chinchiná, Caldas, Colombia, (9)Cornell University/ School of Integrative Plant Sciences/ Plant Pathology and Plant Microbe Biology Section, Santa Barbara, CA
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 (Denoeud *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. eugenioides*. 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*. 2013) prior to genome annotation and high through put diversity variant calling. We also generated transcriptome assemblies for *C. arabica,* using different tissues and conditions of biotic and abiotic stress. Transcriptome assemblies have been aligned to the genome assemblies for validation.
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 Bioinfomatics and Computational Biology (Bios).
This abstract will have an extended time (40 min) and will be 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).

W172: Coffee Genomics
Insights from the Genome of the Major Coffee Insect Pest Worldwide: The Coffee Berry Borer
Lucio Navarro[1], Flor Edith Acevedo[1], Jonathan Nuñez[1], Erick Hernandez[2], Rita Daniela Fernandez-Medina[3], Claudia Carareto[4], Pablo Benavides[1] and Stuart Jeffrey[5], (1)Centro Nacional de Investigaciones de Cafe, CENICAFE, Chinchiná, Colombia, (2)Sao Paulo State University, Sao Jose do Rio Preto, Brazil, (3)Escola Nacional de Saúde Pública, Fundação Oswaldo Cruz, Rio de Janeiro, Brazil, (4)UNESP - Universidade Estatual Paulista, São José do Rio Preto, SP, Brazil, (5)Purdue University, West Lafayette, IN
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 will be 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.

W173: Coffee Genomics
Towards GWAS and Genomic Prediction in Coffee: Development and Validation of a 26K SNP Chip for *Coffea Canephora*
Alan C. Andrade[1], Orzenil B. da Silva Junior[2], Fernanda de A. Carneiro[3], Pierre Marraccini[4,5] and Dario Grattapaglia[2], (1)Embrapa Café/INOVACAFÉ - UFLA, Lavras-MG, Brazil, (2)Embrapa Recursos Genéticos e Biotecnologia, Brasília-DF, Brazil, (3)Graduate Program on Plant Biotechnology - UFLA, Lavras-MG, Brazil, (4)Embrapa Recursos Genéticos e Biotecnologia, Brasilia-DF, Brazil, (5)CIRAD 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 24,073 SNPs (94.6%) successfully converted out of the 25,456 SNPs on the array. 20,982 markers(87.1%) were scored as providing high-resolution genotypic data in a set of 800 individuals of a breeding population in which 19,586 (81.4%) were polymorphic and 1,396 (5.8%) were monomorphic. The remaining set of 3,091 (12.8%) successfully converted markers were of lower accuracy in the studied sample and may require additional cluster analysis to proper biological interpretation, *i.e.* targeting CNVs. 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.

W174: Coffee Genomics

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

Luis Felipe V. Ferrão, University of São Paulo (ESALQ/USP), PIRACICABA, Brazil, Romario G. Ferrão, Instituto Capixaba de Pesquisa,Assistência Técnica e Extensão Rural, Vitoria, Brazil, Maria A. G. Ferrão, INCAPER/EMBRAPA CAFÉ, Vitória-ES, Brazil, Aymbire Fonseca, Instituto Capixaba de Pesquisa e Assistência Técnica e Extensão Rural / Embrapa Café, Vitoria, Brazil, Matthew Stephens, University of Chicago, Chicago, IL and Antonio Augusto Franco Garcia, University of São Paulo (ESALQ/USP), Piracicaba - SP, Brazil

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.

W175: Coffee Genomics

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

Kassahun Tesfaye, Addis Ababa University, Addis Ababa, Ethiopia and Feyera Senbeta, College 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.

W176: Coffee Genomics

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

Allen Van Deynze[1], Amanda M. Hulse-Kemp[1], Michael C. Schatz[2], Jason Chin[3], Jay Ruskey[4], Dario Cantu[1] and Juan F. Medrano[1], (1)University of California, Davis, CA, (2)Johns Hopkins University, Baltimore, MD, (3)Pacific Biosciences, Menlo Park, CA, (4)Good 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.

W177: Comparative Genomics

A Comparative Genomic Analysis of Plant Reproductive phasiRNAs