

International Coffee Genomics Network (ICGN) Report Coffee Genomics Workshop

Plant and Animal Genome (PAG-31) Meeting,
San Diego, California, January 12-17, 2024

https://plan.core-apps.com/pag_2024/event/0121205505a9f11e6727dfd2f4a883fc

Abstracts

Whole genome assembly, annotation, and alignment tools to tap *Coffea* diversity for climate change adaptation of high quality *Coffea arabica*

Marcela Yepes¹, Carlos Ernesto Maldonado², Aleksey Zimin^{3,4}, Carmenza E. Góngora², Claudia Flórez², Alvaro Gaitán² and Herb Aldwinckle¹, ⁽¹⁾ Cornell University/ School of Integrative Plant Sciences/ Plant Pathology and Plant Microbe Biology Section, Geneva, NY, ⁽²⁾ FNC/CENICAFE Colombian National Coffee Research Center, Chinchiná, Colombia, ⁽³⁾ University of Maryland, College Park, MD, ⁽⁴⁾ Johns Hopkins University, Department of Biomedical Engineering, Baltimore, MD

There is pressing need to characterize and utilize the germplasm of the world's most valuable export commodity: Coffee. There are more than 130 species in the *Coffea* genus, but the current varieties cultivated to produce coffee — a market valued at \$200 billion — have very little genetic diversity. To address disease and climate pressures that threaten production, we are generating ultra-high quality genome assemblies for *Coffea arabica*, the most cultivated coffee species in the world and the only tetraploid species of the genus *Coffea*.

Ultra-high quality, telomere-to-telomere (T2T), haplotype-resolved, gapless genome assemblies are significantly more complete, correct, and contiguous, and allow one to conduct population scale studies to accelerate not only characterization of diversity in *Coffea* but its utilization into *C. arabica* breeding programs. Long-read sequencing is starting to make an impact in population sequencing for coffee due to reduced costs and significant improvements in data quality and workflows. We are using long reads generated with both Pacific Biosciences (PacBio) High Fidelity (HiFi) Revio, and ultralong reads generated with Oxford Nanopore PromethION. In collaboration with Cantata Bio, we are generating chromosome scaffolded, haplotype resolved genome assemblies for our target *C. arabica* genotypes and for several *Coffea* species, and are optimizing whole genome alignment methods to study genome evolution and diversity not only for protein-coding genes but more importantly for identifying genome duplications, chromosome rearrangements, and the widespread structural variations (SVs). This information is needed to advance our understanding of intra- and inter-specific structural variation in coffee, genomic function, and sequence evolution. EviAnn (short for Evidence-based annotation), a novel evidence-based eukaryotic gene annotation software (Zimin, *et al.* 2023) that outperforms current state of the art machine learning genome annotation packages like BREAKER3, MAKER2, and FINDER, while utilizing less computer time, is being used to annotate our coffee genomes.

Diploid and Polyploid resolved genome assemblies

Sierra McWilson

SMcWilson@cantatabio.com

Director Sales Cantata Bio

Tom Swales

TSwale@cantatabio.com

Senior Manager Business Operations Cantata Bio

Cantata Bio launched June 6, 2022, with the mission of enabling researchers to address the world's most challenging scientific questions, from human disease to agricultural sustainability, using leading-edge multi-dimensional next generation sequencing (NGS) technologies. The company is a result of the merger between Dovetail Genomics, the industry leader in advanced proximity ligation genomic solutions, and Arc Bio, which develops novel, proprietary metagenomic tools for accurate and sensitive microbial profiling. Committed to delivering the most innovative NGS-based solutions, Cantata Bio comprises three business units, Epigenetics & Genome Structure, Microbial Profiling, and Genetic Analysis Solutions.

Cantata Bio offers polyploid assembly and has completed 11 haplotype-resolved polyploid assemblies with additional projects in process. The success in generating polyploid assemblies represents a significant advancement in genomics. By refining assembly methods and prioritizing data integrity, we have constructed sub-genome and haplotype-resolved assemblies that surpass previous pseudo-haploid assemblies in completeness and contiguity. These results not only showcase the potential of cutting-edge genomic tools in understanding complex plant genomes but also set a new standard for future polyploid genome assembly projects.

Cantata Bio's polyploid assembly analysis paves the way for similar in-depth functional genomics studies, cultivar improvement, and precise genetic manipulation. The provided comprehensive dataset serves as a critical resource for the scientific community, enabling researchers to uncover genetic underpinnings of traits like disease resistance and drought tolerance, which are crucial for the enhancement of plants or crops worldwide, including coffee.

Persephone: an interactive visualization platform for browsing across multiple genomes

Maxim Troukhan, Ph.D.

mtroukhan@persephonesoft.com

Persephone Software, LLC

Persephone® is a multi-genome browser designed to display multiple genomes side-by-side in a fluid, animated fashion to facilitate genome visualization and exploration. Major advantages are fast data retrieval, speed, interactivity, and ability to handle large amounts of data in real time. The browser can display diverse data including genetic maps, genomic sequences, gene models, synteny, QTLs, variants, RNA-seq, read alignments, and other features mapped to a genome. The ability to show multiple chromosomes on one screen allows interactive exploring syntenic

relationships between genomes to facilitate connecting orthologous genes. The users of the web application (<https://web.persephonesoft.com>) can add their own data and run comparative analyses, such as aligning entire chromosomes, in real-time. The high data visualization rate facilitates navigation across multiple genomes. Examples of the tool functionality will be presented using multiple species, including coffee.

Genomic-assisted breeding of *Coffea canephora* for climate-smart coffee

Luis Felipe Ferrao, PhD

lferrao@ufl.edu

University of Florida, Gainesville

Coffee is a universal beverage that drives a multi-industry market on a global basis. Accelerated climate changes threaten the sustainability of coffee production. Herein, we propose the implementation of genomic-assisted breeding for climate-smart coffee in *Coffea canephora*. Important factors to consider when developing climate-resilient coffee crops are increased adaptation to higher temperatures and resilience to biotic and abiotic stresses. To this end, we evaluated two populations, over multiple harvests, and under drought conditions to dissect the genetic architecture of yield, disease resistance, and quality-related traits. By integrating genome-wide association studies and diallelic analyses, our contribution is four-fold: (i) First, we identified a set of molecular markers with major effects associated with key traits; (ii) We demonstrated the relevance of non-additive gene actions and projected hybrid vigor when genotypes from geographically different botanical groups are crossed; (iii) We reported genetic parameters that could drive future breeding decisions; and finally (iv) We plan to use molecular breeding to maximize genetic gains. Altogether, this work is a blueprint for how quantitative genetics and genomics can assist coffee breeding and support the supply chain in the face of the current global changes.

High-Throughput Sequencing of the 16S rRNA Gene during Coffee Fermentation after Inoculation of Bacterial Starter Cultures in Northern Peru to Improve Coffee Quality

Martha S. Calderon^{1,2*}, Danilo E. Bustamante^{1,2*}, Samia L. Fernandez-Güimac¹, Jhordy Perez¹, Jani E. Mendoza¹, Jose Isaias Barboza Saldaña¹

¹Instituto de Investigación para el Desarrollo Sustentable de Ceja de Selva, Universidad Nacional Toribio Rodríguez de Mendoza, Chachapoyas, Amazonas, Perú

² Instituto de Investigación en Ingeniería Ambiental, Facultad de Ingeniería Civil y Ambiental, Universidad Nacional Toribio Rodríguez de Mendoza, Chachapoyas, Amazonas, Perú

Peru is the ninth largest coffee producer and the top exporter of organic coffee worldwide. In its production process, fermentation is an important stage where microorganisms break down the mucilaginous substrate of coffee beans and shape the smell and flavor of the final coffee drink. The lack of innovation in the fermentation process is a problem that affects the coffee-growing regions of northeastern Peru. In this sense, starter cultures (i.e., a preparation that contains large quantities of microorganisms) emerge as a technological tool to standardize the fermentation process and ensure good and constant coffee quality. Prototypes of starter cultures were developed

using combinations of three strains of bacteria with pectin-degrading activity (*Stenotrophomonas* sp.1, *Stenotrophomonas* sp.2 and *Lysinibacillus* sp.). In this study, the microbiota of spontaneous coffee fermentation after inoculation of potential starter cultures was evaluated by high-throughput sequencing of the 16S rRNA gene. The study was carried out on a coffee farm in Amazonas that produces coffees ranked with 78 points. A 2x4 factorial experimental design was used and each treatment was composed of a mixture of bacteria. Furthermore, to evaluate the effect of fermentation time, two durations were used, 14 and 36 hours (h) after the fermentation started. In total, 27 samples were obtained, including a control sampled at the beginning of fermentation (0 hours). The control samples collected at 14 h and 36 h were mainly composed of *Leuconostoc* and *Lactobacillus*, respectively, and in both times *Bacillus* spp. was absent. Regarding the samples with starter culture, there was a higher abundance of *Leuconostoc*, *Lactobacillus*, *Stenotrophomonas/Gluconobacter* at 14 h, while *Leuconostoc*, *Acetobacter*, *Komagataeibacter* and *Acetobacteraceae* were predominant at 36 h. Within each treatment (i.e., type of starter culture) a variation in the relative abundance of *Acetobacter* and *Stenotrophomona* was observed. Sensory analysis showed an increase in the cupping score from 2.5 to 5 points, including significant differences based on the combination of the three and two bacteria (*Stenotrophomona* sp.2 + *Lysinibacillus* sp.) Finally, variation between fermentation times appears to have a greater impact on species composition than the types of starter cultures used.

Keywords: *Coffea arabica*, coffee fermentation, metabarcoding 16S rRNA.

Funding: METACAFÉ 2.0 N° PE501081184-2022-PROCIENCIA-CONCYTEC

Unveiling metabolome diversity in coffee varieties

Jeffrey Chu, PhD

jeff.chu@metware-bio.com

Metware Biotechnology Inc

Mingjuan Zhang, Metware Biotechnology Inc.

Fanguang Hu, Institute of Tropical and Subtropical Cash Crops, Yunnan Academy of Agricultural Sciences

Coffee is one of the most consumed non-alcoholic beverages in the world and an understanding of its nutritional values and differences between varieties and throughout its processing is important. A key technology to achieve this understanding is metabolomics. Using an innovative metabolomics approach, the Widely-Targeted Metabolomics, and one of the largest plant metabolite databases, Metware Biotechnology is able to gain a detailed view of the metabolic profiles from different varieties and processing methods. Using this technology, we have looked at the metabolic profile of coffee berries from four *Coffea arabica* variety Bourbon sub-varieties differing in berry color. The four sub-varieties showed diverse nutritional aspects: those with yellow fruits showing higher coumarins, lignans, and alkaloids; those with orange fruits showing the highest organic acids and terpenoids; purple fruits showing the highest content in phenolics and flavonoids. With insights from metabolomics, coffee can be marketed with related health benefits specific to the variety in addition to their unique aroma and flavor.