

## POSTER - DNA AND GENOMICS

### **GENOME-WIDE ASSOCIATION STUDY FOR LIPIDS USING A REFERENCE GENOME OF COFFEA ARABICA**

*Herison Victor Lima Muniz (herison.muniz@hotmail.com)*

*Caroline Ariyoshi (carolineariyoshi@hotmail.com)*

*Rafaelle Vecchia Ferreira (rafaelle.ferreira@gmail.com)*

*Mariane Silva Felicio (marianesfelicio@gmail.com)*

*Bruna Silvestre Rodrigues Da Silva (brunasilvestrerodrigues@hotmail.com)*

*Alexandre Rossi Paschoal (paschoal@utfpr.edu.br)*

*Luiz Filipe Protasio Pereira (filipe.pereira@embrapa.br)*

Lipids are key coffee compounds playing an important role in coffee bean development, contributing to the flavor and aroma of coffee beverages. *Coffea arabica* is an allotetraploid species ( $4n = 2x = 44$ ) that belongs to the genus *Coffea*, which originated from the natural cross between two diploid species, *Coffea canephora* and *Coffea eugenioides*. Genome-wide association study (GWAS) is a powerful approach used to identify associations between genetic variants and a trait of interest. In a previous GWAS of lipids in coffee, the genotyping by sequencing (GBS) data was aligned on its ancestor *C. canephora* reference genome. In this study, our objectives were to identify quantitative trait nucleotide (QTNs) and genes related to lipids in coffee from GBS data previously aligned to the *C. arabica* genome, from the accession ET039, provided by Arabica Coffee Genome Consortium, in which 159,000 SNP

markers were identified in 159 coffee accessions (Felicio; Rodrigues, 2018, data not published). The GBS data from 106 accessions of *C. arabica* was filtered using TASSEL 5.2.53 software for minimum allele frequency (MAF > 0.05) and call rate (CR > 0.8). To adjust the models was performed the Principal Component Analysis (PCA) for population structure using TASSEL 5.2.53 and kinship matrix (matrix K) using the mrMLM package. To identify the associated genomic regions, we performed GWAS using six models: multi-locus random-SNP-effect mixed linear model (mrMLM), a fast mrMLM multilocus mixed linear model (FASTmrMLM), FAST multi-locus random-SNP-effect EMMA (FASTmrEMMA), Iterative modified-Sure Independence Screening EM-Bayesian LASSO (ISIS EM-BLASSO), polygenic-background-control-based least angle regression plus empirical Bayes (pLARmEB) from the mrMLM package, and Fixed and random model Circulating Probability Unification (FarmCPU) from the GAPIT package, using R software for both packages. The critical value to select SNPs were: p-value = 0.01, 0.01, 0.005, 0.01 for mrMLM, FASTmrMLM, FASTmrEMMA and ISIS EM-BLASSO, respectively, with critical LOD score = 3 for all models from the mrMLM package, and p-value = 0.05 for FarmCPU. In our study, after the quality control MAF and CR of GBS data, 11,136 SNPs from 106 accessions were selected for GWAS, a number approximately four times greater than observed in a previous study from Sant'Ana et al. (2018). GWAS identified 11 QTNs related to lipids, approximately six more than in the previous study. The use of GBS data aligned to the *C. arabica* genome demonstrated an increase in SNP identification, enabling the GWAS to detect QTNs potentially involved in the lipid trait. Verification of the associated genomic regions in the new alignment on the genome of *C. arabica* ET39 is underway, to identify possible genes involved in lipids metabolism.