

Non coding RNAs in *Coffea canephora* genome: Identification by similarity.

Samara M. C. de Lemos¹, Alexandre R. Paschoal¹, Douglas S. Domingues^{1,2}

¹Graduation Program in Bioinformatics – PPGBIOINFO – Universidade Tecnológica Federal do Paraná, Cornélio Procopio; ²Department of Botany, Instituto de Biociências, Universidade Estadual Paulista, Rio Claro

Non-coding RNAs (ncRNAs) are an important component of genomes and transcriptomes of eukaryotes. They are preferentially generated in intronic and intergenic regions of the genome. ncRNAs can comprise microRNAs (miRNAs), transporter RNAs (tRNAs), small nucleolar RNAs (snoRNAs) and small nuclear RNAs (snRNAs). Fundamental processes for plant physiology and development, such as flowering and fruit ripening are regulated by ncRNAs, which motivates the identification of ncRNAs in plant genomes. Coffee genus has at least a hundred species and some of them has an important role in agricultural industry. Brazil is the leading producer and second largest consumer market of coffee. Recently, the genome of Roubusta coffee (*Coffea canephora*) became available for analysis. This species accounts for one third of the world production. This study aimed the identification of ncRNAs in the newly sequenced *C. canephora* genome. Intergenic and intronic regions were extracted using a pipeline developed with *Shell* and *Perl* scripts. We BLASTed these regions against ncRNA public data from 31 plant species available in ENSEMBL Plants (version 32) and applied a *Perl* script to filter hits above 80% of identity and coverage. Using these criteria, we identified 47 hits in *C. canephora* introns, 13 miRNAs, 25 tRNAs, 7 snoRNAs and 2 snRNAs. From intergenic regions we obtained 525 alignments. They were classified as 273 snRNAs, 231 rRNAs, 14 ncRNAs and 7 miRNAs. The 3 species that presented most hits in coffee were *O. sativa* (rice), *S. tuberosum* (potato) and *S. lycopersicum* (tomato). These preliminary results provide a starting point for the characterization of a genomic component that is a protagonist in several agronomical traits.

Acknowledgement: Arabica Coffee Genome Consortium.