

# Genome-wide Identification of Metabolic gene Clusters in *Coffea canephora*

Samara Mireza Correia de Lemos<sup>1</sup>, Suzana Ivamoto Suzuki<sup>2</sup>, Luiz Filipe Pereira<sup>3</sup>, Alexandre Paschoal<sup>4</sup>, Douglas Domingues<sup>1</sup>

<sup>1</sup> Sao Paulo State University, Institute of Biosciences, Rio Claro, Brazil.

<sup>2</sup> State University of Londrina, Londrina, Brazil.

<sup>3</sup> Brazilian Agricultural Research Corporation (Embrapa Café), Londrina, Brazil.

<sup>4</sup> Federal University of Technology – Paraná, Cornélio Procopio, Brazil.

Corresponding author: Samara Lemos - e-mail: samara.lemos@unesp.br

## INTRODUCTION

Metabolic gene clusters (MGCs) are groups of physically proximal genes that are responsible for the biosynthesis of specific natural products. They are already known in bacteria, yeasts, filamentous fungi, mammals and plants. MGCs can be related to primary and/or specialized metabolic pathways. Specialized metabolites from plants are extensively used as medicines, drugs, pigments, cosmetics, nutraceuticals and agrochemicals. In coffee, several compounds responsible for the complexity of coffee flavor and aroma are specialized metabolites. We then hypothesized that MGCs differentially expressed in coffee beans might contribute to biochemical composition of the beverage.

## METHODS

*C. canephora* genome sequences

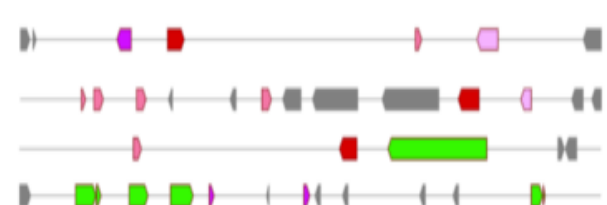
9 seed RNASeq libraries in *C. canephora*

Coffee Genome Hub  
a genomics and genetics resource for coffee



planti  
SMASH

Metabolic Gene Clusters



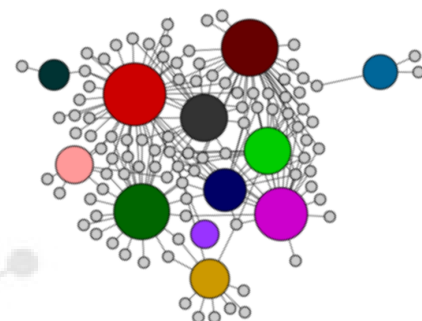
IDEAMEX

- edgeR
- limma-Voom
- DESeq2
- NOISeq

*C. canephora* DEG

33 *C. canephora* RNASeq libraries (seeds and leaves)

Modules of coexpressed genes

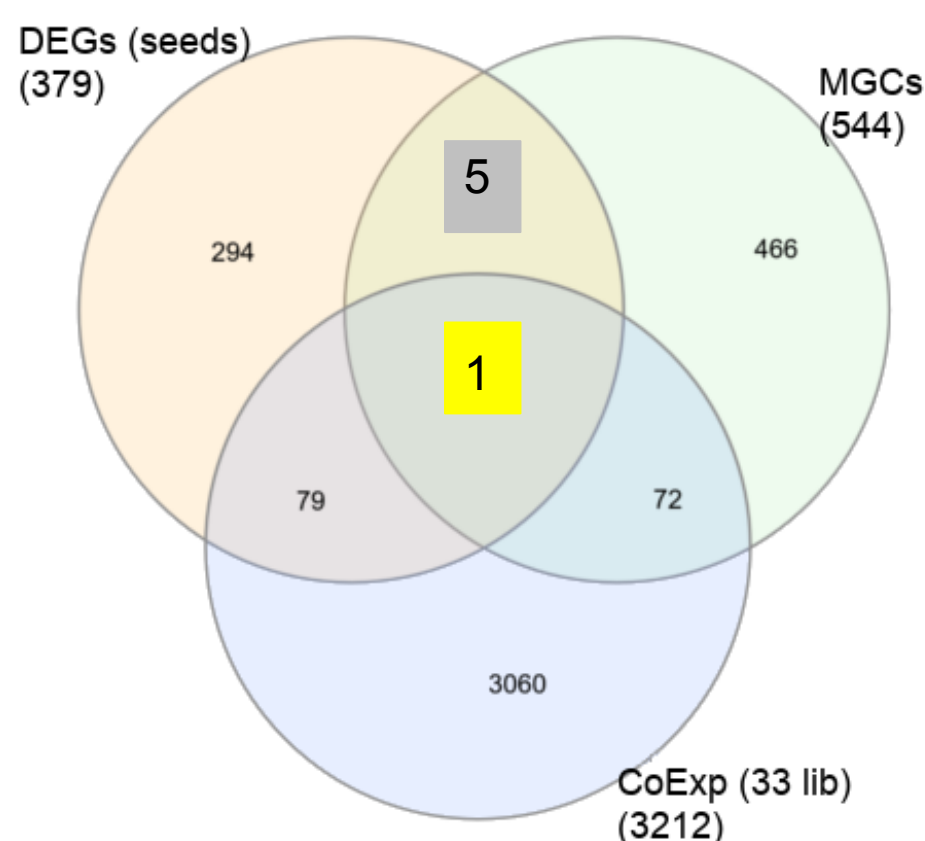


CEMiTool

webCEMiTool

## RESULTS

- 46 MGCs identified in *Coffea canephora* assembled chromosomes: saccharides (19), putative (12), terpenes (6), alkaloids (5), lignans (3), polyketides (1);
- 6 gene coexpression modules in 33 RNAseq libraries;
- 379 Differentially expressed genes in *Coffea canephora* seeds;
- 73 genes were found in 35 MGCs and 5 coexpression modules;



Tabersonine 16-hydroxylase  
MGC 7, chr 3 (terpene)  
Module 1

Gene	Metabolic Gene Cluster
Serine carboxypeptidase-like 2	7, chr1 (saccharide)
Hypothetical protein	2, chr2 (saccharide)
Phospho-2-dehydro-3-deoxyheptonate aldolase 1	1, chr10 (saccharide)
Squalene monooxygenase	2, chr10 (saccharide)
F-box protein At1g70590	3, chr 11 (alkaloid)

## CONCLUSION

The analysis of MGCs in *C. canephora* genome with coexpression and DEGs, provides novel insights into the potential regulatory mechanisms and helps the identification of key genes for further studies on the molecular biology of metabolism in coffee trees.

## REFERENCES

- Denoeud F, Carretero-Paulet L, Dereeper A, et al. The coffee genome provides insight into the convergent evolution of caffeine biosynthesis. *Science*. 2014;345(6201):1181-1184.
- Kautsar SA, Suarez Duran HG, Blin K, Osbourn A, Medema MH. plantiSMASH: automated identification, annotation and expression analysis of plant biosynthetic gene clusters. *Nucleic Acids Res*. 2017;45(W1):W55-W63.
- Jiménez-Jacinto V, Sanchez-Flores A, Vega-Alvarado L. Integrative Differential Expression Analysis for Multiple EXperiments (IDEAMEX): A Web Server Tool for Integrated RNA-Seq Data Analysis. *Front Genet*. 2019;10:279.
- Cardozo LE, Russo PST, Gomes-Correia B, et al. webCEMiTool: Co-expression Modular Analysis Made Easy. *Front Genet*. 2019;10:146.
- Anna K Stavrinides, Stéphane Dussert, Marie-Christine Combes, et al. Seed comparative genomics in three coffee species identify desiccation tolerance mechanisms in intermediate seeds, *Journal of Experimental Botany*, Volume 71, Issue 4, 7 February 2020, Pages 1418–1433.

## ACKNOWLEDGEMENTS

