

Private alleles in Brazilian *Coffea arabica* cultivars

Caroline Ariyoshi¹ (ariyoshi.caroline@gmail.com), Gustavo H. Sera¹, Thierry Leroy², Joseph Mouen³, Luiz F. P. Pereira^{4/1} IDR-Paraná, Londrina, Brazil; ² CIRAD, Montpellier, France; ³ IRAD, Yaoundé, Camaroon; ⁴ EMBRAPA Café, Brasília, Brazil

Introduction

To improve disease resistance and climate resilience in *Coffea arabica*, breeding programs have used introgression from diploid *Coffea* species. The Sarchimor and Catucaí groups originated from *C. canephora* via the Timor Hybrid 832/2 and Icatu, respectively, while *C. liberica* introgression occurred through the BA-10 hybrid. Private alleles (PAs) are alleles unique to a single population or cultivar within a broader genetic pool and can serve as genetic signatures. This study aimed to identify PAs in arabica cultivars derived from diploid species introgression.

Material/Methods

- GBS data from 134 pure Arabica accessions were analyzed in comparison with accessions from the Sarchimor group (IAPAR 59, IPR 99, IPR 104, IPR 107, Tupi IAC 1669-33), the Catucaí group (IPR 102, IPR 103), and cultivars containing BA-10 introgression (IPR 101, IPR 105).
- Tags (80 bp) per genotype were extracted using TASSEL GBS v2 and mapped to the ET39 HiFi reference genome with Bowtie2.
- For PAs identification, a custom Python script selected tags present only in cultivars carrying introgressions from diploid species.
- Gene prediction based on PAs mapping was performed with AUGUSTUS, followed by functional annotation using eggNOG-mapper v2.

Table 1. Some examples of PAs inserted within genes with functional annotations that indicate potential for validation and application in biotechnological tools to support Arabica breeding.

PAs ID	chr	functional annotation	cultivars
1c_46613483	1c	Plant disease resistance NBS-LRR	Tupi
1c_48292936	1c	HEAT SHOCK PROTEIN 26	IAPAR 59; IPR 99; Tupi
1c_48952470	1c	Plant disease resistance NBS-LRR	IAPAR 59; IPR 99; Tupi
1c_57498443	1c	PIF3 (PHYTOCHROME INTERACTING FACTOR 3)	IAPAR 59; IPR 104; IPR 107; Tupi
2c_15384447	2c	Thaumatin family/pathogenesis-related group 5 (PR5)	IAPAR 59
3c_21475673	3c	HEAT-INDUCIBLE TRANSCRIPTION REPRESSOR	IPR 101; IPR 105
3c_32767923	3c	Mitogen-activated protein (MAP) kinase	IPR 101; IPR 105
3c_7802665	3c	Disease resistance protein, plants	IPR 101; IPR 105
3c_8407809	3c	Disease resistance protein, plants	IPR 101; IPR 105
4c_43813887	4c	S-type anion channel (response to water deprivation; response to salt stress)	IAPAR 59; IPR 104; IPR 107; Tupi

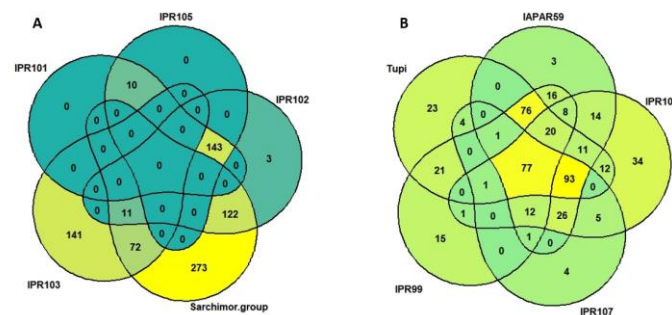


Figure 1. Distribution of 776 PAs identified. **A** – Among cultivars from the Sarchimor group, Catucaí cultivars (IPR 102, 103), and cultivars introgressed with *C. liberica* (IPR 101, 105). **B** – Among Sarchimor cultivars (Tupi IAC 1669-33, IAPAR 59, IPR 99, 104, 107).

Results/Discussion

- A total of 776 PAs, including both cultivar-specific and shared variants, were identified (Figure 1).
- PAs from the Sarchimor group showed higher concentration on chromosome 1, whereas those from the Catucaí group and cultivars with BA-10 introgression were mainly located on chromosome 3 (Figure 2).
- Functional annotation revealed PAs inserted within genes involved in responses to biotic and abiotic stresses, circadian rhythm regulation, photoperiodism, and the development of vegetative and reproductive organs.
- Both cultivar-specific and shared PAs, based on their functional annotations, exhibit potential for validation and application in breeding programs as molecular markers linked to agronomic traits, as exemplified in Table 1.

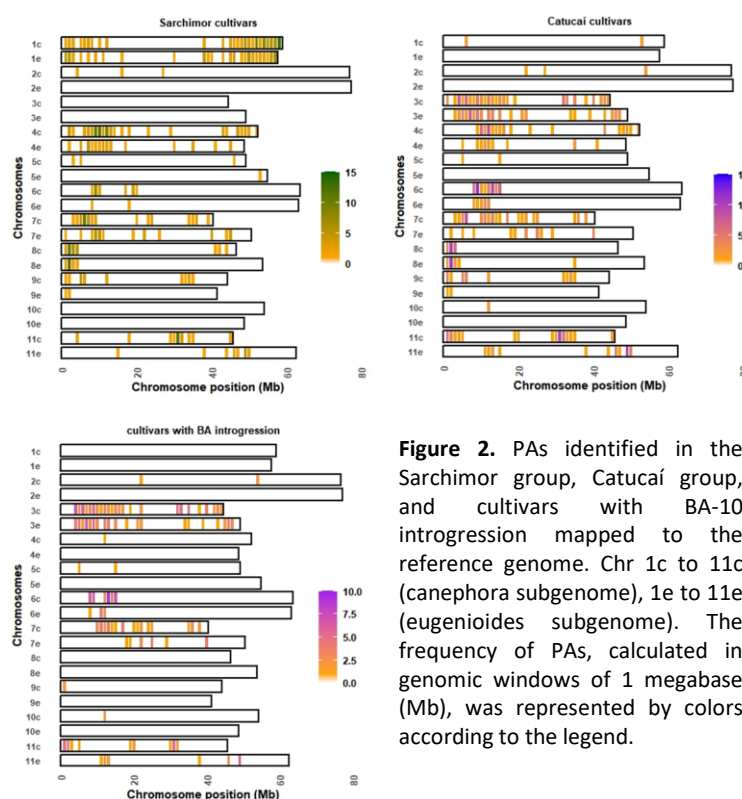


Figure 2. PAs identified in the Sarchimor group, Catucaí group, and cultivars with BA-10 introgression mapped to the reference genome. Chr 1c to 11c (canephora subgenome), 1e to 11e (eugenioides subgenome). The frequency of PAs, calculated in genomic windows of 1 megabase (Mb), was represented by colors according to the legend.