WHITE PAPER

Introduction to MolBreeding's Coffee 40K Panel

Executive Summary

The Coffee 40K Panel developed by MolBreeding represents a significant advancement in the field of coffee genetics and breeding. This panel is designed to facilitate the identification and differentiation of coffee varieties through SNP genotyping. The goal of this paper is to outline the purpose, methodology, and potential applications of the Coffee 40K Panel, highlighting its importance in coffee research and breeding programs.

Introduction

Coffee is one of the most widely consumed beverages globally, and its cultivation is integral to many economies. With increased demand for high-quality coffee and the challenges posed by climate change and diseases, there is a pressing need for innovative breeding solutions. The Coffee 40K Panel is a tool aimed at enhancing the genetic understanding and breeding of coffee plants, particularly focusing on the *Coffea arabica* and *Coffea canephora* species.

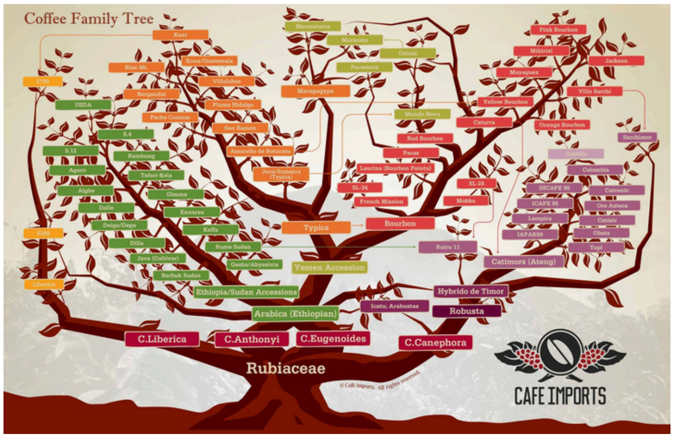


Figure 1. Coffee family tree
Resource: Coffee Plant Varietals

Methodology

Sample Collection

The Coffee 40K Panel employs a diverse range of samples sourced from multiple origins:

- Germplasm Pool: A total of 76 samples, primarily from the Yunnan Coffee cultivar collection in China and public collections at the Tropical Academic of Science Institute.
- *C. canephora* Germplasm Pool: A total of 6 samples, hypothesized to be pure lines closely related to *C. canephora*, as indicated by GWAS results.
- NCBI Reference Samples: A total of 84 samples, which include supplementary information from published studies and represent a wide genetic diversity of coffee. For reference, see Scalabrin et al. 2024 PubMed.

Whole genome sequencing data were obtained for these samples and utilized for SNP selection.

SNP Selection

During the SNP selection process, we applied specific criteria to form the Coffee 40K Panel - v1.0. In total, 166 samples were used to call SNPs, from which we selected SNPs based on the following criteria to ensure high quality:

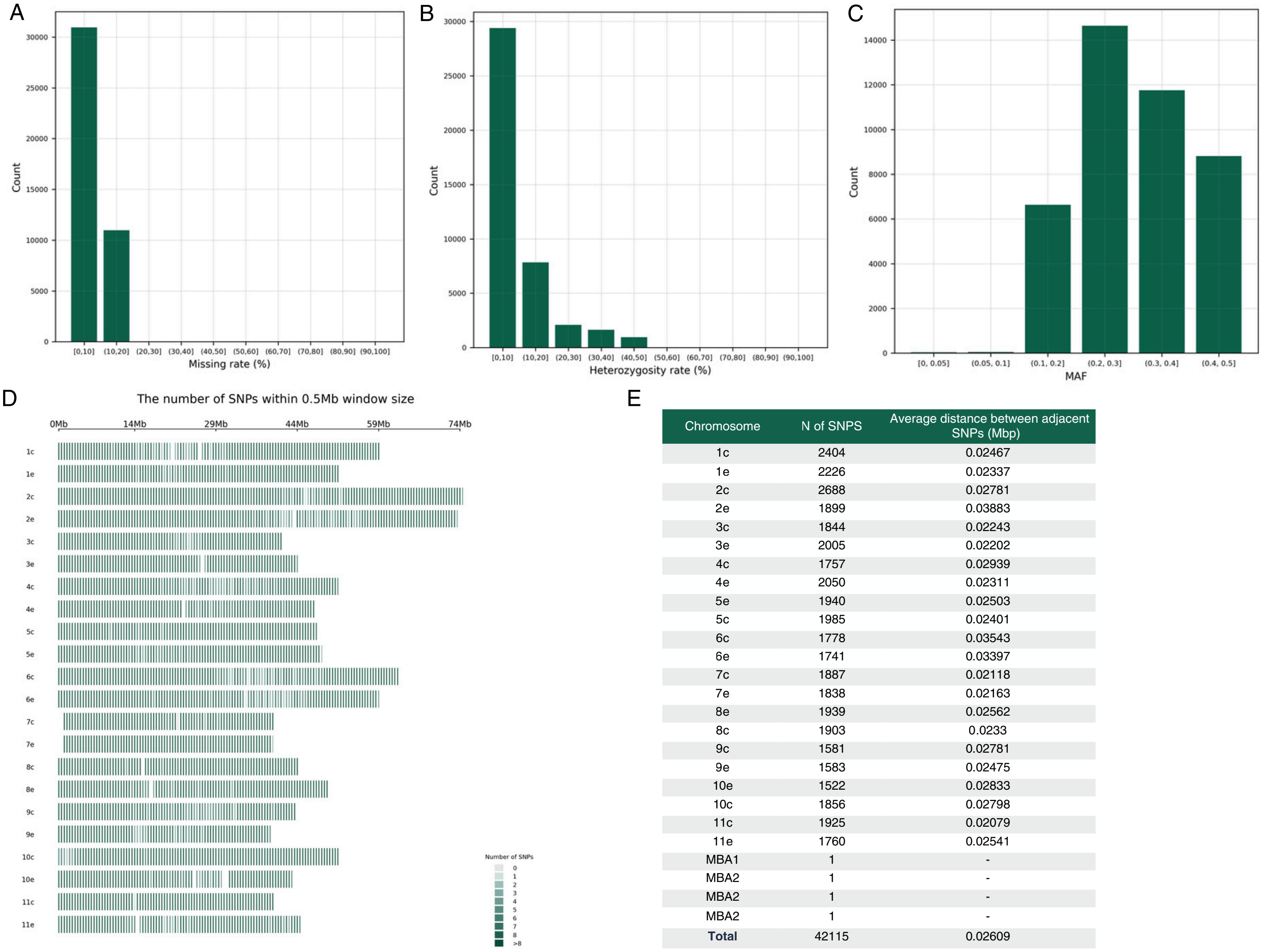
- Missing rate < 10%. (Fig. 2A)
- Heterozygosity ≤ 50%. (Fig.2B)
- Minor Allele Frequency (MAF) preferred ≥ 0.35. (Fig.2C)
- Additional SNPs included to fill gaps, ensuring a uniform distribution across the reference genome.

In addition to the selection of the 40K SNPs, we incorporated 45 publicly available SNPs from WRC into the collection. These SNPs can also differentiate 23 varieties from Latin America.

GOALS

1. SNP Selection for Coffee Varietie Identification: Identify specific SNPs used for genotyping different coffee varieties, enabling precise genomic analysis.

2. Variety Differentiation: Evaluate the effectiveness of the Coffee 40K panel in distinguishing between different coffee species and hybrids, including but not limited to *Coffea arabica*, *Coffea canephora*, and their hybrids.



Chromosome	N of SNPS	Average distance between adjacent SNPs (Mbp)
1c	2404	0.02467
1e	2226	0.02337
2c	2688	0.02781
2e	1899	0.03883
3c	1844	0.02243
3e	2005	0.02202
4c	1757	0.02939
4e	2050	0.02311
5e	1940	0.02503
5c	1985	0.02401
6c	1778	0.03543
6e	1741	0.03397
7c	1887	0.02118
7e	1838	0.02163
8e	1939	0.02562
8c	1903	0.0233
9c	1581	0.02781
9e	1583	0.02475
10e	1522	0.02833
10c	1856	0.02798
11c	1925	0.02079
11e	1760	0.02541
MBA1	1	-
MBA2	1	-
MBA2	1	-
MBA2	1	-
Total	42115	0.02609

Figure 2. Data-performance summary of the Coffee 40K panel.
(A) Missing-rate distribution. (B) Heterozygosity profile. (C) Minor-allele-frequency (MAF). (D) Distribution of target genotyping sites on chromosomes. (E) Number of SNP markers on each chromosome.

Panel Design

Panel Name: GenoBaits® Coffee 40K Panel
Panel ID: PHR0365_Coa40K_V1.0
Reference genome: *Coffea_arabica*_IGA_Cara_2.4 (*C. arabica*, Tetraploid)
Total SNPs: 42,115 high-quality SNPs were retained, evenly distributed across all chromosomes of the reference genome with a mean inter-marker distance of 0.26 Mb (Fig. 2D–F).

Data Analysis

The Coffee 40K Panel employs advanced statistical methods to analyze SNP data, allowing researchers to:

- Conduct Genome-Wide Association Studies (GWAS) to identify genetic markers associated with desirable traits.
- Perform Principal Component Analysis (PCA) to visualize genetic relationships among coffee varieties.

Population Analysis based on the 40K SNPs selected

To further evaluate the 40K SNPs selected in this coffee panel, we utilized publicly available 166 samples (NCBI accessions) to construct a phylogenetic tree and analyze the relationships among these samples, including *C. arabica, C. canephora, C. eugenioides, Red Bourbon, SL28 (Bourbon), Typica, and C. arabica* × *C. canephora* introgression lines. PCA analysis was performed based on the selected 40K SNPs (Fig. 3B, 3D, 3F). To compare the tree with established coffee family trees (Fig.1), we excluded non classification samples and conducted additional analyses to demonstrate concordance with published data (Fig. 3A, 3C,3E).

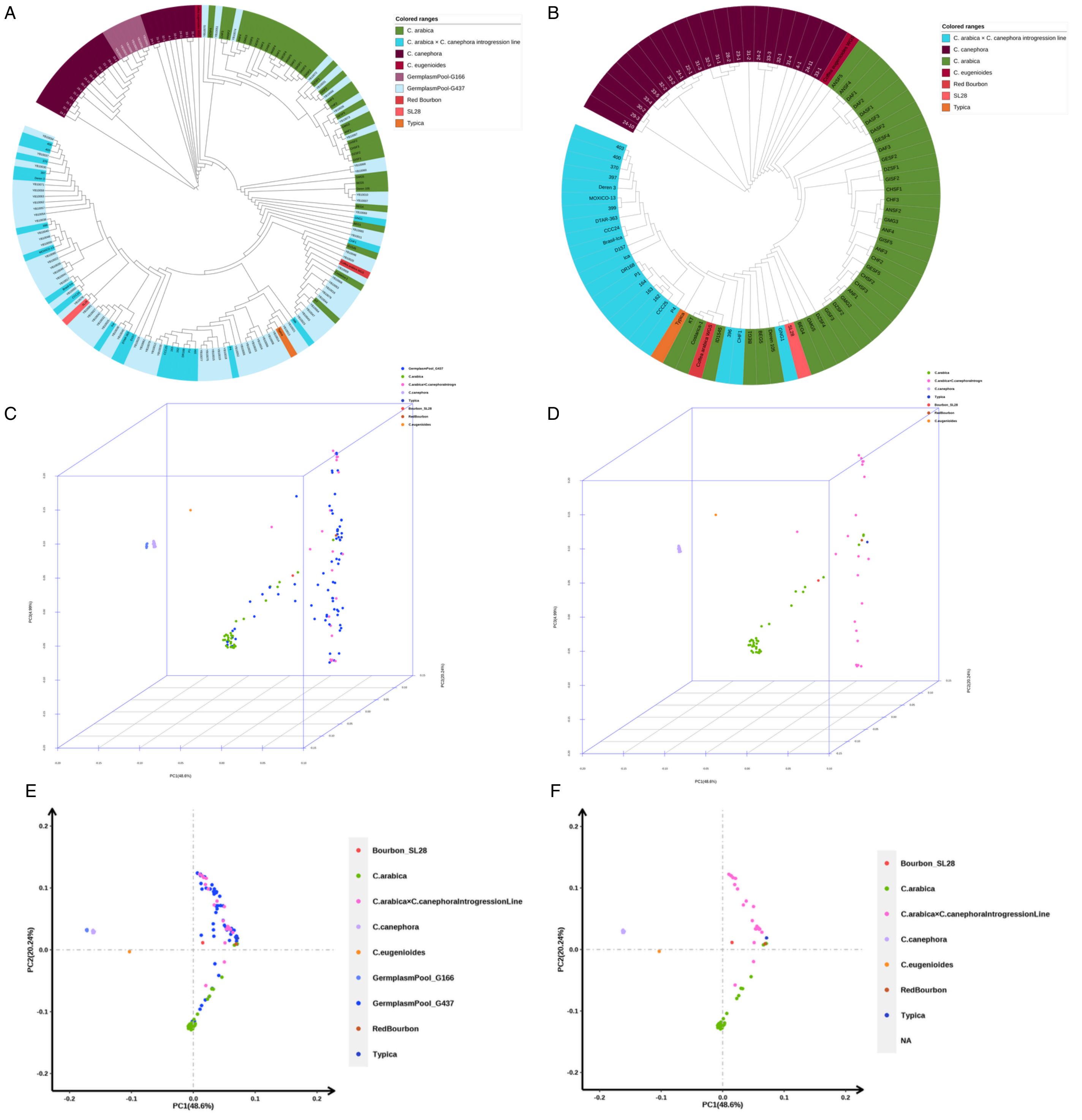


Figure 3. Population Analysis based on the 40K SNPs selected

(A) The phylogenetic tree of 166 coffee samples including *C.arabica, C.canephora, C.eugenioides, Red Bourbon*, SL28 (*Bourbon*), *Typica*, *C.arabica×C.canephora* Introgression Line and and two non-classified pools (GermplasmPool_G166 and GermplasmPool_G437). (B) The phylogenetic tree of 84 coffee samples after removal of non-classified samples. (C,E) The PCA analysis of 166 coffee samples. (D,F) The PCA analysis of 84 samples.

Applications

The applications of the Coffee 40K Panel are vast and can significantly impact the coffee industry:

1. Breeding Programs: By identifying genetic markers associated with quality, yield, and disease resistance, breeders can make informed decisions, accelerating the development of superior coffee varieties.

2. Genetic Research: The panel provides a robust framework for researchers to explore the genetic diversity of coffee, aiding in the understanding of traits related to flavor, aroma, and growth characteristics.

3. Sustainability: With climate change affecting coffee production, the panel can help in developing varieties that are more resilient to environmental stresses, thereby ensuring sustainable coffee farming practices.

Conclusion

The Coffee 40K Panel by MolBreeding represents a pioneering step in coffee genetics and breeding. By facilitating the identification of SNPs and enhancing the differentiation of coffee varieties, this panel holds the potential to transform coffee research and breeding practices. As the coffee industry continues to evolve, tools like the Coffee 40K Panel will play a crucial role in meeting the challenges ahead, ensuring that coffee remains a beloved beverage for generations to come.

Call to Action

We invite coffee researchers, breeders, and industry stakeholders to engage with us in utilizing the Coffee 40K Panel for advancing coffee genetics. Together, we can ensure a sustainable and high-quality future for coffee cultivation.

Reference

1. Scalabrin S, Magris G, Liva M, et al. A chromosome-scale assembly reveals chromosomal aberrations and exchanges generating genetic diversity in *Coffea arabica* germplasm. Nat Commun. 2024;15(1):463.
2. Mekbib Y, Tesfaye K, Dong X, Saina JK, Hu GW, Wang QF. Whole-genome resequencing of *Coffea arabica* L. (Rubiaceae) genotypes identify SNP and unravels distinct groups showing a strong geographical pattern. *BMC Plant Biol*. 2022;22(1):69.
3. Scalabrin, S., Toniutti, L., Di Gaspero, G. et al. A single polyploidization event at the origin of the tetraploid genome of *Coffea arabica* is responsible for the extremely low genetic variation in wild and cultivated germplasm. Sci Rep 10, 4642 (2020).
4. Huang L, et al. Resequencing 93 accessions of coffee unveils independent and parallel selection during Coffea species divergence. Plant Mol. Biol. 2020;103:51–61.

MolBreeding Biotechnology | cs@molbreeding.com

Americas | 10 Pasteur, Suite 150,Irvine, CA 92618 United States | +1-515-508-9174
EMEA | Unit 10, West Point Business Park, West Road, Harlow, Essex, CM20 2BU United Kingdom | +44 (0) 1279 940 983
Asia | 538B, 5/F, Building 1W, 1 Science Park W Ave, Science Park, Pak Shek Kok, Hong Kong SAR, China | + (852) 5130-1856