bioRxiv preprint doi: https://doi.org/10.1101/2023.06.15.544964; this version posted June 15, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Characterization of the genetic composition and establishment of a core collection for the INERA Robusta coffee (*Coffea canephora*) field genebank from the Democratic Republic of the Congo

Lauren Verleysen[1,2], Robrecht Bollen[1,3], Jean-Léon Kambale[4], Tshimi Ebele[5], Benjamin Ntumba Katshela[3], Jonas Depecker[1,3], Valérie Poncet[6], Dieu-Merci Assumani[5], Filip Vandelook[1,3], Piet Stoffelen[3], Olivier Honnay[2,7] and Tom Ruttink[1,8*]

[1] Division of Ecology, Evolution and Biodiversity Conservation, Faculty of Sciences, KU Leuven, Leuven, Belgium

[2] Plant Sciences Unit, Flanders Research Institute for Agriculture, Fisheries and Food, Melle, Belgium

[3] Meise Botanic Garden, Meise, Belgium

[4] Université de Kisangani, Kisangani, Democratic Republic of the Congo

[5] Institut National des Etudes et Recherches Agronomique, Yangambi, Democratic Republic of the Congo

[6] UMR Diversité, adaptation, développement des plantes (DIADE), Univ. Montpellier IRD, CIRAD, Montpellier, France

[7] Plant Institute, KU Leuven, Leuven, Belgium

[8] Department of Plant Biotechnology and Bioinformatics, Faculty of Sciences, Ghent University, Ghent, Belgium

*** Correspondence:**
Corresponding Tom Ruttink
tom.ruttink@ilvo.vlaanderen.be

Keywords: *Coffea canephora*, HiPlex amplicon sequencing, genetic structure, core collection, *ex-situ* conservation, crop wild relatives, Democratic Republic of the Congo.

bioRxiv preprint doi: https://doi.org/10.1101/2023.06.15.544964; this version posted June 15, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

ABSTRACT

Cultivation of Robusta coffee is likely to gain importance because of its high disease resistance and climate envelope, but substantial effort must be made to improve its cupping quality and other agronomic traits. Robusta coffee genetic resources conserved in field genebanks can play an important role in this context, but many of them are limited and poorly characterized. In this study, we aimed to characterize the genetic composition of the historically important but until recently neglected INERA Coffee Collection in Yangambi (Democratic Republic of the Congo) and to propose a Robusta core collection. We combined genome-wide GBS marker sets with a novel versatile multiplex amplicon sequencing (HiPlex) screening assay, for the genetic screening of the INERA Coffee Collection and created tools for future genetic screening. A set of 730 coffee shrubs in the collection were genotyped at 86 HiPlex loci and 263 unique genetic fingerprints were identified. Representative fingerprints were compared to 'Lula' cultivars, 'Luki' cultivars, Congolese subgroup A, and wild genotypes growing in the rainforest of the Yangambi region. The majority of the unique genetic fingerprints were identified as 'Lula' cultivars but also four Congolese subgroup A and nine wild genotypes were found in the collection. Twenty-five percent of the unique genetic fingerprints had a hybrid identity which could result from seed-based propagation of the collection or natural cross-pollination. Moreover, eight unique genetic fingerprints were frequently identified as a parent and these may correspond to the ancient selected mother plants, or direct descendants thereof, of which seed material was widely distributed from 1930 to 1960 for the creation of coffee plantations. Finally, to delineate core collections we tested the maximization strategy and genetic distance method for the INERA Coffee Collection, and proposed i) a core collection conserving the loci diversity of the collection, *i.e.* 100 accessions and ii) a core collection holding the highest genetic distances between the accessions, *i.e.* 10 accessions. Our study provides a method for the genetic characterization of Robusta coffee collections in general and contributes to the revaluation and exploitation of the Robusta coffee genetic resources in the Democratic Republic of the Congo in particular.