BMC Plant Biology

RESEARCH

Open Access

Uncovering the historic genetic diversity of Ugandan Robusta coffee germplasm: a foundation for core collection design

Mildred Julian Nakanwagi[1,2,6*], Peter Wasswa[1], Mathieu Gonin[2,3,4], Remi Auguste[3,4], Tram Bao Vi[3,4], Pascal Musoli[2], Pierre Marraccini[3], Catherine Kiwuka[5], Jeninah Karungi[1], Isaac Onziga Dramadri[1], Mildred Ochwo Ssemakula[1], Claude Patrick Millet[4,7], Ali Milton[2], Geofrey Arinaitwe[2], Godfrey Sseremba[2], Boris Delahaie[3,4†] and Valerie Poncet[4†]

Abstract

Background Uganda is a primary centre of diversity for *Coffea canephora* (Robusta coffee), a crop that underpins both the national coffee sector and global breeding programs. The National Coffee Research Institute (NaCORI) of Uganda conserves more than 2058 accessions that were collected from different agroecological zones of the country. However, their genetic diversity has been only partly characterized, limiting their effective use in conservation and breeding.

Results A total of 909 *C. canephora* accessions from the NaCORI collection was genotyped, using 236 KASPar SNP markers. To investigate genetic structure and diversity we combined multivariate analyses, genetic diversity and differentiation metrics as well as hierarchical clustering. Genetic clusters were then spatially mapped according to the agroecological regions of origin across Uganda. Using reference genotypes from known wild Ugandan forests and major African genetic groups we observed that accessions from NaCORI were majorly related to the genetic group O from Uganda (and to a lesser extent Congolese group E). Despite the low differentiation among regional subpopulations of origin, reflecting gene flow driven by historical dissemination and admixture within the collection, the cultivated accessions were separated into six distinct clusters using Ward's hierarchical method. From these genetic clusters, two core collections were designed using Core Hunter 3. The first comprised 109 accessions with high allelic coverage (CV = 0.981), while the second comprised 328 accessions that, in addition to allelic coverage, optimized the representation of both geographical origin and genetic clusters. These corresponded to 13% and 42% of the entire studied collection, respectively. The core collection of 328 accessions was defined as the GWAS population for downstream association studies. Including representative wild Ugandan accessions in the analyses reveals that some forests may harbour unique diversity absent from the cultivated germplasm.

Conclusion Characterizing this diversity and designing representative core and GWAS collections are essential to optimize resource management and enable the targeted development of climate-resilient cultivars. The multi-

†Boris Delahaie and Valerie Poncet contributed equally to this work.

*Correspondence:
Mildred Julian Nakanwagi
mildredjuliannakanwagi@gmail.com

Full list of author information is available at the end of the article

regional origin of our core collections revealed the influence of ecological and socio-cultural factors on germplasm diversity. Our results also suggest that including wild accessions from Ugandan forests could greatly enhance the cultivated genetic pool.

Keywords *Coffea canephora*, *Ex situ* conservation, Geographical sources, Early culture histories, Crop wild relatives, Core collection, KASPar markers

Introduction

As climate change accelerates, it exacerbates episodes of rising temperatures, erratic rainfall, and shifting pest dynamics [1]. These environmental changes disrupt the ecosystems and threaten both cultivated crops and their wild relatives by making it increasingly difficult for them to cope with the rapid changes [2]. The challenge is greater for perennial crops whose cropping cycle is longer with slower genetic turnover, making them less responsive to abrupt shifts in climate and ecological conditions [3, 4]. As a result, their inability to adapt swiftly increases the risk of crop failure, reduces long-term productivity, and jeopardizes the stability of food systems that rely on these species [5]. Furthermore, it accelerates biodiversity erosion among landraces and wild relatives, ultimately compromising the genetic diversity which is essential for sustaining resilient agricultural systems [6]. This loss of genetic diversity undermines the resilience of agricultural systems, reducing breeder's capacity to develop crops that can withstand future environmental stresses thereby threatening the long-term sustainability of global food production [7]. Importantly, the sustainability of crop production in the face of climate change hinges on the conservation and strategic use of plant genetic material, hence the urgent need to preserve them [8]. Conservation efforts must not only safeguard diversity but also ensure it remains accessible and relevant to those who need it for example the breeders, and farmers. In situ and *ex situ* conservations are the major two complementary strategies that play a critical role in safeguarding valuable genetic resources that have essential traits for climate resilience and future crop improvement [9]. In situ conservation focuses on maintaining genetic resources within their natural habitats, allowing populations to evolve through environmental pressures under natural selection thereby preserving adaptive traits [10]. *Ex situ* conservation offers stability and accessibility but may not fully capture the dynamic range of diversity shaped by environmental change [8]. While *ex situ* collections have served as stable repositories for crop diversity over years [11], they are neither immune to climate-related risks. They are further challenged by extreme weather events, infrastructure vulnerabilities, and shifting ecological baselines which compromise the integrity and representativeness of stored materials [12]. In cases where diversity is threatened, wild populations offer a critical fallback for breeders to tap into adaptive traits that may be absent from cultivated stocks. This can ultimately address the issue of genetic bottlenecks that occur when released or cultivated varieties lose diversity due to continuous selection [13, 14].

Africa, as the centre of origin for many *Coffea* species, hosts a wide range of locally adapted species across diverse ecological zones [15]. Among them, *Coffea canephora* Pierre ex A. Froehner exemplifies this complex dynamic facing in situ germplasm conservation. *C. canephora* is native from the evergreen forests with a wide ecological distribution spanning from Central, through West and East Africa [15]. This wide distribution not only highlights the ecological breadth but also the multifaced pressures causing vulnerability of wild coffee genetic resources and conservation complexity [16]. Following this, there has been an increasing recognition of the relevance of genetic diversity as strategic for the evaluation, selection, and development of resilient varieties tailored to diverse agroecological conditions [17]. Successive genetic studies on wild *C. canephora* material have identified eight distinct native genetic groups (A, B, C, D, E, O, G, and R) originating from diverse native ecological environments [18–21]. These native source populations have contributed to varying extents, to the genetic makeup of the cultivated populations [19, 22]. While Uganda's native populations have been assigned to group O [23], the origin and diversity of its cultivated accessions are more complex and remains insufficiently documented.

Although informal cultivation of *C. canephora* from probable local wild plants dates back several centuries before 1900 [24], in regions such as Gabon, Guinea and Uganda, its spread and commercial cultivation began only in the late 19th century [25]. This was rapidly adopted due to its resistance to coffee rust (*Hemileia vastatrix*) which had devastated *C. arabica* and *C. liberica* plantations especially in Southeast Asia [26]. In response to this threat, Java breeding centre in Indonesia received *C. canephora* plants from different countries including Democratic Republic of Congo (DRC) and Uganda [25]. Ugandan accessions (*C. canephora* var. Ugandae (Cramer) A. Chev) were characterised with great resistance to the coffee berry borer [25]. More germplasm collections were subsequently conducted in Sankuru (Congo Basin regions), Gabon, Angola and Uganda for the second time. These were incorporated into the Java breeding programs in Indonesia thereby forming the backbone

of *ex situ* conservation and breeding strategies. Germplasm selected from Java was later re-distributed to other breeding centres in Asia [22] and Africa. One of the first reintroductions was made to DRC particularly to the INERA Yangambi station which played a central role in assembling and characterizing native *C. canephora* populations from the Congo basin starting in the 1930s [26, 27]. Other breeding centres were established in Madagascar in 1930, and these incorporated both local and introduced germplasm, including accessions from Congo and Uganda [26]. In Ivory Coast formal breeding activities began in the 1940s, sourcing Robusta material from Java, Congo, Uganda, and other African regions, to develop high-yielding varieties suited to West African conditions [18]. The national coffee breeding program of Uganda was formalized since 1916 when structured selection efforts started, building on the country's rich native diversity and a few earlier contributions from international germplasm exchange (Java) [21, 28]. This global movement of genetic material underscores the interconnectedness of breeding programs and the foundational role of Ugandan germplasm in shaping Robusta cultivation worldwide.

The crop has a long tradition in Uganda where it is embedded in different social and cultural norms. Hundreds of years before the 19th century, indigenous people harvested coffee seeds from the wild forest such as Mabira, Budongo, Kasai and Kibale forests, planting them around homesteads forming one of the earliest known cases of Robusta cultivation globally [21, 24]. During this time coffee was majorly for cultural rituals such as sacred offerings in Buganda where coffee beans cemented social bonds (a tradition which persists in the traditional weddings) and welcomed guests [28]. The species spread via human migration and intertribal trade, reaching the Eastern, Western, and Southwestern regions. Later this traditionally cultivated Robusta was classified into two morphotypes: Erecta, with upright stems and vigorous growth [28], and Nganda, a spreading bush type. To date, the distinctions between the two remain obscure which underscores the need to validate whether initial morphological classifications align with genetic differentiation [23, 29]. Historical records indicate that these morphotypes were widely distributed to farmers in seed form between 1920 and 1970. Following this, a series of more systematic breeding and selection programs conducted by the Coffee Research Station (now NaCORI) led to the release of 8 elite clonal varieties in 1980 [29, 30]. These were selected for their agronomic performance and formed the backbone of Uganda's Robusta plantations for decades. However, the outbreak of Coffee Wilt Disease (CWD) in 1993, caused by *Fusarium xylarioides* devastated Robusta production, prompting an urgent breeding response [29]. A nationwide collection of survivor plants from severely wilt affected farmer fields was launched in the major coffee production parts of the country. These were screened and out of them resistant genotypes were identified resulting in selection of 7 Kawanda Resistant (KR) lines in 2009. Three additional KR Robusta coffee clones were released in 2017. These lines KR1 through KR10 were bred majorly for resistance to CWD, and improved yield [29]. Currently, climate change forecasts for East Africa predict a mean temperature increase of 1.5–2 °C by mid-century and more erratic rainfall patterns [31]. Under such conditions, even the KR clones bred primarily against CWD may succumb to heat stress and emergent pests and diseases that are expanding their elevation range [32]. This calls for broadening the cultivated genetic base given the fact that reliance on a restricted number of parental genotypes accelerates vulnerability to both known and emerging threats [33]. To that cause, NaCORI launched efforts to collect and conserve additional germplasm from on-farm, on-farm trees near forests [34] as well as from wild *C. canephora* populations [23]. To date, over 2,058 cultivated accessions are assembled, but the lack of adequate characterization and evaluation of this germplasm constrains both conservation planning and breeding progress like in any other *ex situ* gene bank [9].

Previous studies of *C. canephora* genetic diversity in Uganda have primarily focused on wild Robusta populations, mainly sampling a limited number of forests and relatively small sample size [21, 23]. When cultivated accessions were included (n = 52), sampling was largely restricted to the central region, thus failing to capture the full breadth of genetic variation across the country. Consequently, the characterization of cultivated germplasm collected nationwide, and its position within the broader spectrum of *C. canephora* diversity, remains unclear. This question is particularly relevant in Uganda where *C. canephora* diversity reflects a complex history of introductions and germplasm collections originating from diverse ecological zones. As a result, NaCORI collections likely represent one of the most genetically diverse repositories of cultivated Robusta compared with most collections that largely reflect local or introduced material [35–39].

Kompetitive Allele-Specific PCR (KASP) assays provide an efficient, highly reproducible, and cost-effective approach for genotyping while maintaining robust resolution of genetic diversity [22, 40, 41]. Here, we applied KASPar markers developed by [22] which have been proven effective for assessing Robusta diversity and tracing the origin of cultivated accessions within the global diversity of the species. These markers enable large-scale genotyping of cultivated germplasm and generate the genetic information required to characterize *ex situ* collections, detect redundant accessions, and identify

under-represented lineages [42]. These molecular data also support the development of well-defined core collections in coffee [22], an effective strategy for maintaining a reduced number of accessions while maximizing allelic diversity and facilitating germplasm evaluation and long-term conservation [43–46]. By applying KASPar markers to a uniquely diverse Ugandan germplasm set collected across major coffee-growing regions [47], this study investigates the genetic diversity and structure of cultivated *C. canephora* within the NaCORI collection to support germplasm conservation and breeding. Specifically, we aimed to: (i) determine the relationship between *C. canephora* germplasm conserved at NaCORI and global diversity across the species' African native distribution; (ii) assess the influence of history dissemination and collection sources on the diversity present in NaCORI's collection; (iii) infer genetic kinship and redundancy among accessions to guide the development of representative core subsets and GWAS populations; and (iv) evaluate the potential contribution of novel genetic diversity from wild *C. canephora* populations to the diversity conserved in the *ex situ* collection.

Materials and methods

Study materials

In Uganda, NaCORI has the mandate to maintain and develop sustainable coffee varieties. Currently the institute harbours a substantial germplasm collection of *C. canephora* in field gene banks. Following a germplasm curation conducted in 2023 the gene bank consisted of 2,058 cultivated accessions of *C. canephora.* The accessions were collected in 1990 from different parts of the country, primarily from farms that had been heavily affected by coffee wilt disease (CWD), as reported in the archives [29]. 80% of these were re-tested for CWD resistance in the laboratory and survivors planted in different fields at Kituza, Mukono in the early 2000.

A total of 909 *C. canephora* accessions from the NaCORI *ex situ* gene bank was genotyped in this study. Accessions were selected to represent each family while minimizing the number of closely related samples within families and to capture the range of geographical origins recorded in the passport data (S1 Table). After removing individuals with missing data and genetically redundant accessions (see section *Quality and redundancy control*), 783 accessions were retained for analysis. Of these, 700 accessions were originally collected from 22 districts distributed across five agroecological zones of Uganda (Fig. 1), 51 lacked district-level collection information (coded as "UKN", unknown Ugandan origin), and 32 were identified as improved varieties introduced to Uganda from Java in the 19th century (coded as "Java"). Embedded in the four Uganda regions, the five source agroecological zones differ in climatic conditions, altitude, soil properties, topography, rainfall distribution, and ecosystem characteristics [48, 49]. The Central Region was further divided into two subzones. The first, Lake Victoria "L. Victoria" comprised of districts such as Kassanda, Kayunga, Rakai, Kiboga, and Mubende, and is characterized by moderate annual rainfall ranging from 1000

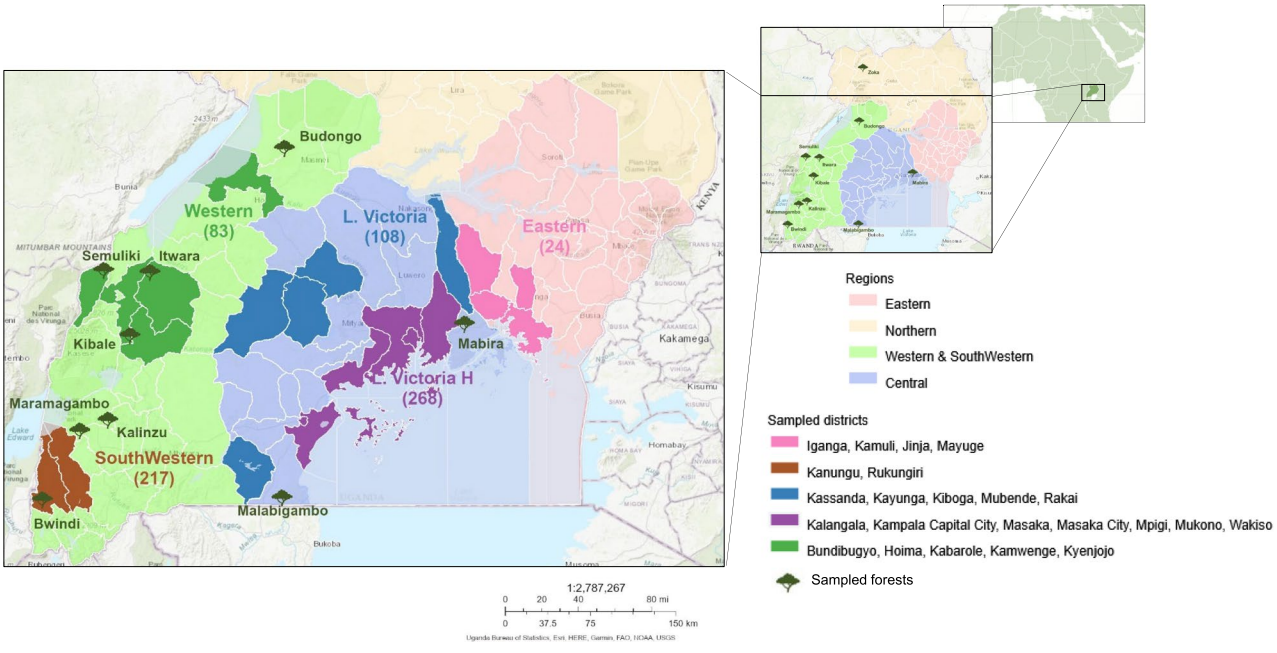

Fig. 1 Geographic distribution of *C. canephora* samples collected across Uganda: spatial spread of 700 coffee samples collected across 22 districts spanning five agroecological zones in Uganda with sample counts in brackets. The majority were from the L. Victoria (5 districts) and L.Victoria H (6 districts) agroecological zones, followed by the Western (5), Eastern (4), and Southwestern regions (2) ones. Sampled forests were represented by tree symbols

to 1200 mm. The other, Lake Victoria adjacent Highlands Zone "L. Victoria H" included Kalangala, Kampala, Masaka, Mpigi, Mukono, and Wakiso, which receive higher rainfall exceeding 1200 mm per year [48, 49]. Other zones included the Eastern (Iganga, Jinja, Kamuli, and Mayuge), Western (Bundibugyo, Hoima, Kabarole, Kamwenge, and Kyenjojo), and Southwestern (Kanungu and Rukungiri) [50]. These represented the landraces, cultivated material, breeders' selections, and Ugandan released varieties (KR lines).

In addition, 50 wild accessions from 10 natural Uganda forests were included in the sample set. All due procedures, including ethical approvals from Uganda National council of Science and Technology (reference number A290ES), were completed prior to sample collection and the acquisition of permits to access the forests. These forests are protected and permission to access and collect study samples was obtained from the National Forestry Authority (NFA; License No. 387) and the Uganda Wildlife Authority (UWA; COD/96/05). The forests sampled were representative of the native distribution range of *C. canephora* across Uganda and included: Kalinzu, Itwara, Kibale, Maramagambo, Malabigambo, Zoka, Semuliki, Bwindi impenetrable, Mabira, and Budongo forests (Fig. 1), each represented by five accessions. In addition, 22 control accessions representing the different established genetic groups were included as reference from previous studies [20, 22]. These included 6 accessions from E group, 6 accessions from A group, 3 accessions from the D group, 2 accessions from R groups, 2 accessions from C group, 2 accessions from O group and one accession from B group.

DNA extraction and KASPar genotyping

Two coffee leaves, picked from the third node below the apical meristem of each accession were dried using 60 g of silica gel. The dried leaves were punched, and four discs were put in the sample plates as provided by LGC Biosearch Technologies. A set of 500 SNP markers evenly distributed across the 11 coffee chromosomes (Figure S1) was selected from the SNP array developed by Mérot-L'Anthoëne et al. [20]. Marker information for these SNPs is available in MoccaDB database (https://moccadb.ird.fr/ [20, 51], to design KASPar-assay and assess genetic diversity. Of these, 261 SNPs had been used in a previous study by [22]. Samples and the designed SNPs primers were submitted to LGC Biosearch Technologies for KASPar genotyping.

Quality and redundancy control

Prior to data analysis, quality control procedures were conducted to filter SNPs and accessions based on proportion of missing data (see workflow in Figure S2). Accessions (n = 7) and markers (n = 4) with more than 35% missing data were identified using AS genomics R-package [52]. Markers were further filtered using a minimum allele frequency threshold (MAF ≥ 0.01), resulting in 236 SNP markers retained for subsequent analyses. Genetically identical individuals were excluded (n = 119) to minimize redundancy. The list of these redundant accessions is provided in S1 Table. From a total of 981 accessions, including reference samples, 855 non-redundant accessions were retained for further analysis (Figure S2).

Genetic diversity assessment

Genetic diversity within the *C. canephora* collection was examined through a series of descriptive analyses across all individuals. Analyses were conducted in R software version 4.4.2 [53]. Genotypic data were imported and handled using the adegenet package version 2.1.10 [54] and converted to hierfstat format [55] for computation of standard diversity metrics. These included estimates of allelic richness, observed heterozygosity (H_O), expected heterozygosity (H_E), and inbreeding coefficient (F_{IS}).

Population structure

Population structure and clustering were analysed using complementary approaches. First, pairwise genetic relationships were assessed using identity-by-state (IBS) distances computed using the SNPrelate package [56]. Principal component analyses (PCA) were performed on two different sample sets: cultivated and reference samples; cultivated samples alone using LEA version. 3.10.2 package [57] and Tidyverse version. 2.0.0 packages in R [58]. IBS values were then transformed into dissimilarities. $Dissimilarity_{ij} = 1 - IBS_{ij}$ where IBS_{ij} is the IBS value between individuals i and j. The resulting distance matrix was subjected to hierarchical clustering using ward's minimum variance (ward.D2) [59]. Clusters were defined by visually inspecting the dendrogram with the primary aim of later examining source populations within each cluster. The resulting hierarchical tree was visualized using the dendextend package [60], with branches color-coded according to cluster membership. Cluster composition was determined and analysed based on sample origins relative to predefined agroecological zones.

Genetic differentiation was quantified for both types of group (i) predefined agroecological zones and (ii) inferred genetic clusters using the unbiased F_{st} estimator implemented in the heirfstat package [61]. Significance of the pairwise distances was evaluated using permutation test with 999 iterations to obtain 95% confidence interval. Analysis of Molecular Variance (AMOVA) was additionally performed to partition genetic variation within and among zones and clusters. These formal tests were used to assess whether agroecological zones or clusters serve

as meaningful grouping factors for structuring genetic variation within the collection.

Core collection development and evaluation

In order to design core collection subsets that preserve overall genetic diversity while ensuring relevance for breeding applications and suitability for downstream association analyses such as Genome-Wide Association Studies (GWAS), we applied a diversity-maximization approach implemented in Core Hunter 3 v3.2.0 [62]. The full initial dataset consisted of 783 unique *C. canephora* accession. These were exclusive of the wild and control accessions used in the study. Fifteen candidate core sets of varying sizes (50–400 accessions in increments of 25) were generated based on the pairwise Euclidean genetic distance criteria [63]. Increments of 25 accessions is a strategy consistent with core collection optimization approaches in which stepwise increases, typically ranging from 20 to 50 accessions, are applied [38, 43, 64]. Furthermore, proportional sampling in core collection development is conventionally set at 5–20% of the total collection size to ensure adequate capture of genetic diversity. In the present study, given a total of 783 accessions, each increment of 25 accessions represents approximately 3.2% of the entire collection. Each core subset was then assessed for genetic diversity and representativeness using the metrics implemented in Core Hunter [62], including Shannon's diversity index (SH), expected heterozygosity (HE), and allele coverage (CV) [65, 66]. Diversity statistics for the core sets were compared with those of the full germplasm collection to evaluate how well each subset captured overall genetic variation. From these comparisons, two core sizes were selected for further evaluation following the classical guideline [43]. The core collections were selected because preliminary comparisons of diversity statistics indicated that this size provided substantially improved allele coverage and geographical representativeness. A fixed group of 10 accessions previously released as NaCORI commercial lines (KR lines) was forced into both selected core sizes to ensure inclusion of breeding-relevant material (when not selected as part). Following the identification of an optimal size range, we refined the selection by incorporating geographical representation, ensuring that accessions from different regions were proportionally included. This resulted into the suggestion of the larger core collection appropriate for GWAS.

Results

Position of NaCORI *C. canephora* germplasm within global diversity panel

To assess the genetic positioning of the NaCORI germplasm within the global diversity of *C. canephora*, a PCA was performed using the studied cultivated accessions from the NaCORI collection together with reference accessions representing established genetic groups [20, 22]. The first two principal components (PCs) clearly separated the major reference groups from the NaCORI materials. Groups A, C, and D representing West African groups formed a well-defined, highly divergent cluster along PC1 positioned far from all NaCORI accessions (Fig. 2A). Groups B, O, E and R were overlapping with NaCORI collection cloud. The NaCORI collection includes accessions historically reported as introductions from Java, whose original African source remains uncertain. In the PCA, Java accessions were broadly dispersed across the NaCORI cluster (Fig. 2B) suggesting mixed or

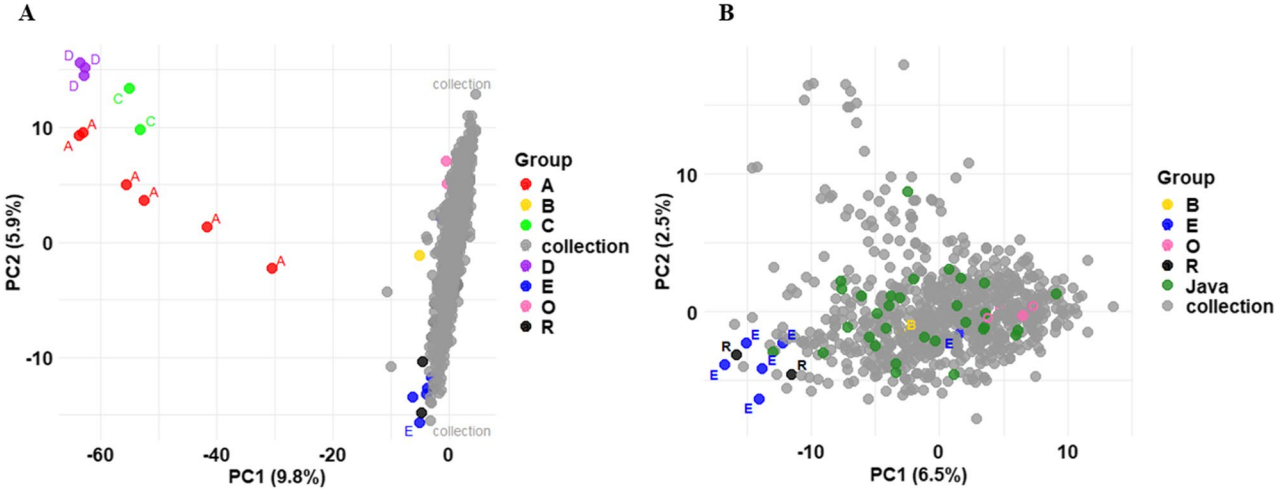

Fig. 2 Principal Component Analysis (PCA) illustrating the position of the NaCORI collection within *C. canephora* global diversity. Relative position of the collection (grey) compared (**A**) to all reference genetic groups as described by [20]; (**B**) to Central (B, E, R) and East African (O) groups; accessions in the collection introduced from Java is shown in green

composite origins within genetic groups B, E, R, O rather than a direct relationship to one group.

Historical dispersal and collection sources effect on NaCORI collection diversity

To examine the relationship between source of origin and observed diversity patterns, a PCA was conducted on cultivated accessions excluding the reference samples (Fig. 3). Accessions formed a broad, continuous distribution along the first two axes with extensive overlap among agroecological zones, and no discrete grouping by origin was observed (Fig. 3A). A few accessions were positioned towards the margins of the distribution, but none formed a distinct cluster. Examination of other axes up to PC4 (Fig. 3B) revealed that some accessions from the Southwestern zone differed from the rest of collection on PC4, indicating a subtle structure.

Pairwise F_{ST} values between zones were overall low, ranging from 0.001 to 0.026 (S2 Table). Although few comparisons were statistically significant ($p < 0.001$), the effect sizes were very small, indicating minimal differentiation among agroecological zones of origin. Only the Southwestern zone showed significant F_{ST} values (F_{ST}>0.001) with all other zones, consistent with its slight displacement on PC4 (Fig. 3B), but divergence remained weak overall. The AMOVA (S3 Table) partitioned 98.51% of the total genetic variance within individuals (attributable to the level of heterozygosity), only 1.13% attributable to differences among zones and 0.37% among individuals within zones. This confirmed that the recorded source zones explain only a small fraction of the overall genetic diversity. Together, these results indicate that variation within the NaCORI collection is largely continuous and only weakly structured by agroecological origin.

Genetic kinship patterns among NaCORI *C. canephora* accessions

Kinship relationships within the NaCORI collection were examined using dendrogram constructed with the IBS distances and Ward's algorithm (Fig. 3C). The dendrogram tree was pruned at the dissimilarity threshold of 0.8 allowing to partition the accessions into six clusters. This threshold was selected to facilitate interpretation, and the resulting groupings were consistent with patterns observed in complementary analyses (e.g. PCA and F_{ST}), indicating that the main conclusions are not driven by this choice. Cluster sizes varied, but all groups contained mixtures of agroecological origins, except cluster 1, that corresponded to the unique agroecological Southwest (Rukungiri district) lineage (Fig. 3C). Some clusters were enriched for particular origins, but most clusters

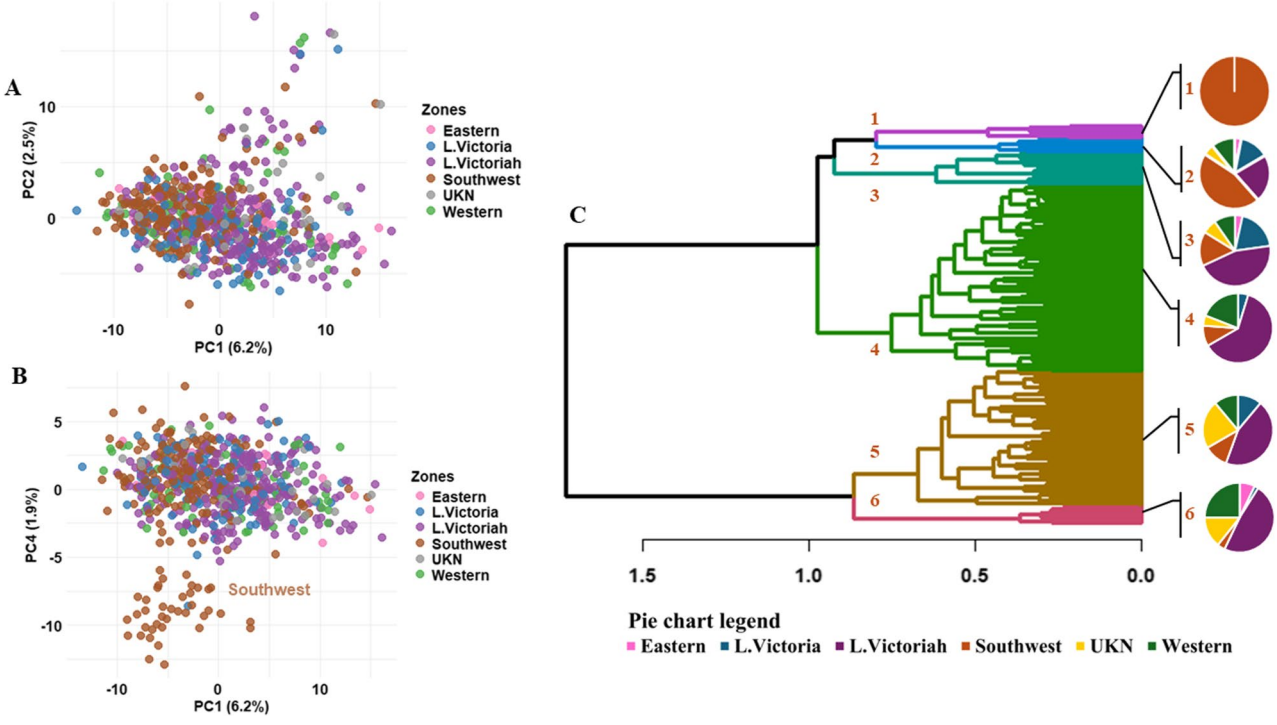

Fig. 3 Genetic characterisation of *C. canephora* accessions from NaCORI collection of Ugandan origin. Principal Component Analysis (PCA) showing the genetic relationships among *C. canephora* accessions from the NaCORI collection (**A**) along PC1 and PC2, and (**B**) along PC1 and PC4; (**C**) dendrogram using Ward clustering using IBS distances with pie charts representing the different agroecological zones of origin for each genetic cluster. Colours on PCAs and pie charts indicate the agroecological of origin as defined on Fig. 1

also included accessions from other zones. Overall, this mixed composition is consistent with the continuous genetic patterns observed in PCA (Fig. 3B) and a slight differentiation of some individuals from the southwest agroecological zone.

The levels of genetic differentiation among these genetic clusters (S4 Table), as estimated by F_{ST} values, ranged from 0.04 to 0.2. High values likely reflect underlying diversity rather than true population subdivision and correspond to the comparison of individuals from the opposite ends of the gradient observed in Fig. 3A. AMOVA likewise attributed only a negligible proportion of variance to differences among clusters, with most of the genetic variation (98.96%) attributed within individual samples. A small proportion of variation (0.96%) was attributed to differences among samples within clusters, while only 0.07% of the total variation was explained by differences between clusters.

Genetic diversity was assessed for the whole collection and across the six clusters using observed heterozygosity (H_O), expected heterozygosity (H_E), and the inbreeding coefficient (F_{IS}). Observed and expected heterozygosity values within clusters were moderate to high and broadly similar across groups (S5 Table). These levels of heterozygosity confirm the presence of substantial standing genetic variation in the collection. F_{IS} values were close to zero and slightly negative, indicating no evidence of inbreeding and a mild excess of heterozygotes, consistent with the outcrossing reproductive system of *C. canephora.* Together, these results and the AMOVA test (S6 Table) show that IBS clustering provides a relative similarity ordering of individuals but does not reveal discrete genetic lineages; instead, the NaCORI germplasm exhibits continuous structure combined with significant genetic diversity.

Core collection evaluation and spatial distribution of NaCORI collection

To determine the optimal core collection size that captures maximum diversity and representativeness, a range of core subset increasing from 50 to 375 at an interval of 25 was evaluated. The results revealed a consistent trend across increasing core sizes (S7 Table). Allele coverage (CV), the proportion of alleles retained from the full collection reached 0.98 by Core3 (109 individuals) and remained constant through Core14 (375 individuals). In contrast, Shannon's diversity index (SH) decreased from 5.92 in Core1 to 5.88 in Core15 (400 individuals), while expected heterozygosity (H_E) dropped from 0.3 to 0.27 over the same range (S7 Table). Two core collections were selected. Core3 comprises 109 accessions and represents 13% of the total collection. This core was designed to maximize allelic diversity. Core12 comprises 328 accessions, representing 42% of the total collection. Core12 was selected to both maximize allelic diversity and ensure broad geographical representation.

To assess the underlying diversity within the core collections and evaluate the representativeness of the two selected ones, a Principal Component Analysis was first conducted (Figs. 4A and S3). The core accessions of Core12 (328 individuals) are broadly interspersed among the non-core accessions. The spread of points along PC1 and PC2 is similar for the core and full collections, spanning approximately the same range on each axis. Similar patterns were observed for Core3 (109 individuals) which is a smaller set and shows a sparse but well-distributed spatial pattern remains (Figure S3).

To further assess whether the core collection reflects the diversity of the full dataset, we compared the distribution of average Euclidean genetic distances (Fig. 4B). The histogram shows that the full collection exhibits a normal distribution with an average distance of approximately 10.7. Most samples were falling between 10 and 11.5. The core collection displays a similar distributional shape to the full set with slightly a narrower spread as compared to the bigger core collection and shifted toward higher pairwise genetic distances (Fig. 4B). This shift indicates that the core collection is well enriched for genetically distinct individuals, maximizing diversity and reducing closely related genotypes. Complementary assessment was done to evaluate the genetic representativeness of the selected core collection relative to the full germplasm set using the distribution of expected (H_E), observed heterozygosity (H_O) and the Shannon diversity index values (SH) (Fig. 4C). The Shannon index and heterozygosity values for the core collection were slightly higher than those of the full collection for both proposed collections (Figs. 4C and S3C). Indeed, the core collections not only captures the overall genetic diversity of the full set but also enhances diversity metrics by retaining more informative and genetically distinct individuals.

To explore the spatial distribution of genetic diversity, core accessions were mapped across Uganda's five studied agroecological zones: Southwestern, Eastern, L Victoria, L Victoria H and Western focusing on districts where the samples were collected (Fig. 4D, Figure S3D). Within each zone accessions were further classified based on their genetic cluster assignment. The results revealed clear differences in cluster composition and geographic representation between the two core collections. Core12 (328 individuals), which included accessions from 19 districts, demonstrated broader spatial and genetic representation than Core3 (109 individuals). Five agroecological zones were represented by accessions from all six genetic clusters, while the Eastern zone comprising three clusters (Fig. 4D). In comparison, Core3 included accessions from only 14 districts of five agroecological, and only one genetic cluster was represented in the Eastern

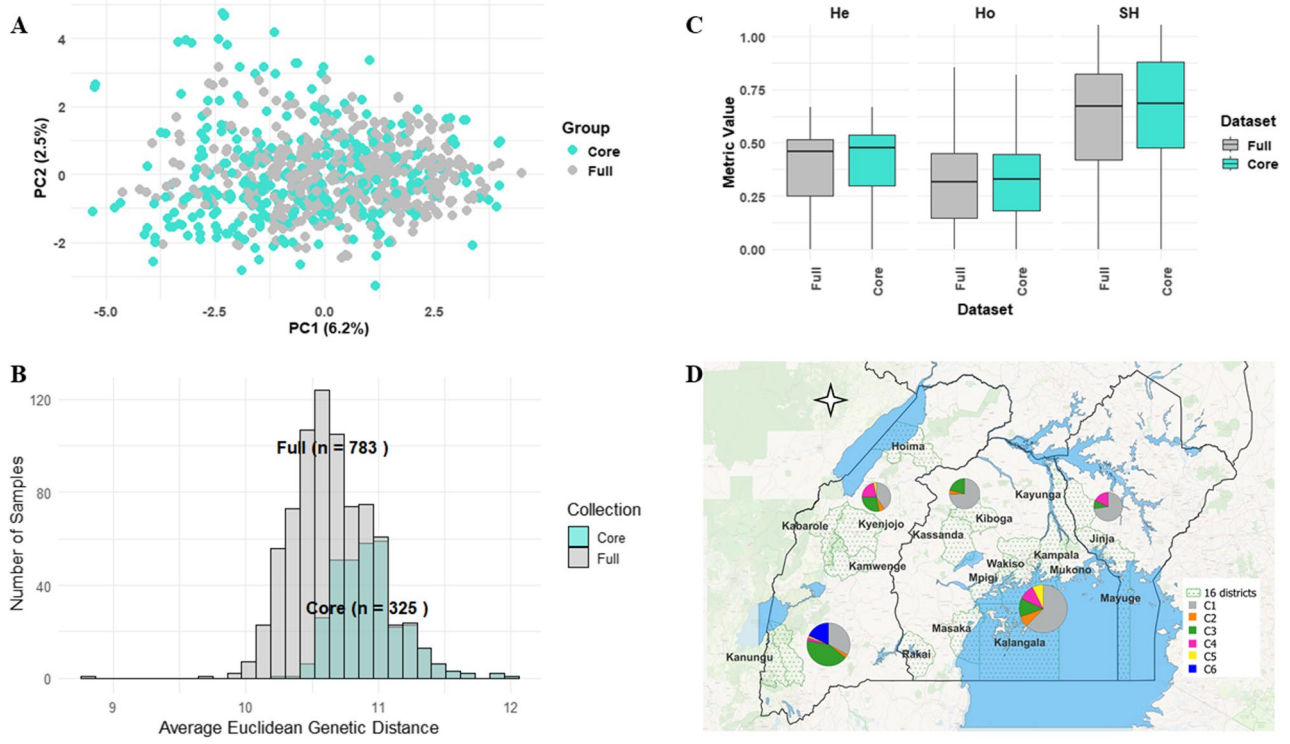

Fig. 4 Characterisation of the core collection Core12 (328 ind) (turquoise) compared to the whole germplasm (grey). (**A**) Principal Component Analysis (PCA) along PC1 and PC2; (**B**) Pairwise Euclidean genetic distances distribution; (**C**) Genetic diversity estimates: expected heterozygosity (HE), observed heterozygosity (H_O) and Shannon index (SH); (**D**) Spatial distribution and genetic cluster composition of the core collection across Uganda's five agroecological regions shown with sampling locations (green dots) and regional genetic cluster proportions (C1–C6) summarized in pie charts

zone (Figure S3D). Overall, Core12 provided a more balanced and comprehensive representation of Uganda's genetic diversity as compared to Core3, both in terms of district coverage and genetic cluster composition.

Potential of Ugandan wild *C. canephora* populations to enhance *ex situ* genetic diversity

To examine how the NaCORI collection relate to wild *C. canephora* diversity found in Uganda, IBS-based Ward clustering was performed on the wild accessions alone and on a combined dataset including both wild and cultivated accessions. On wild accessions, the tree revealed two main groups at higher dissimilarity levels, with subclusters generally corresponding to forest of origin (Fig. 5B). Some forests, such as Itwara and Kalinzu, formed relatively tight subgroups, whereas others (e.g. Semuliki, Mabira) exhibited more internal branching, reflecting differences in the depth of divergence among wild populations.

The clustering analysis incorporating both cultivated and wild accessions (Fig. 5A) revealed the presence of a distinct divergent cluster (C6) that was not observed in the dendrogram constructed using only cultivated accessions (Fig. 3C). While some wild accessions from Semuliki, Malabigambo, Mabira, Maramagambo, Kalinzu and Bwindi forests were interspersed throughout the dendrogram alongside cultivated materials, others, originating from the Kibale, Itawara, Budongo, and Zoka forests, formed a clearly defined and well-separated cluster, cluster 6, indicating substantial genetic divergence from the cultivated gene pool (Fig. 5A).

The dendrogram results were supported by the pairwise F_{ST} analysis (S8 Table), which revealed that Cluster 6 consistently showed strong and statistically significant differentiation from the other clusters ($p < 0.001$; F_{ST} ranging from 0.154 to 0.68). This confirms its genetic uniqueness within the germplasm and highlights the distinctiveness of the wild lineage.

Discussions

Genetic position of NaCORI germplasm within global *C. canephora* diversity

The genetic relationship among global reference groups identified two geographically distinct genetic clusters: the Guinean group, which includes accessions from West Africa, and the Congolese group, comprising populations from Central and East Africa. These clusters and their sub-groups align with those reported in previous studies [18, 20, 22] thereby providing a robust global reference genetic framework. Within the Central and East African lineage, the NaCORI germplasm encompasses much of the genetic diversity observed in reference group

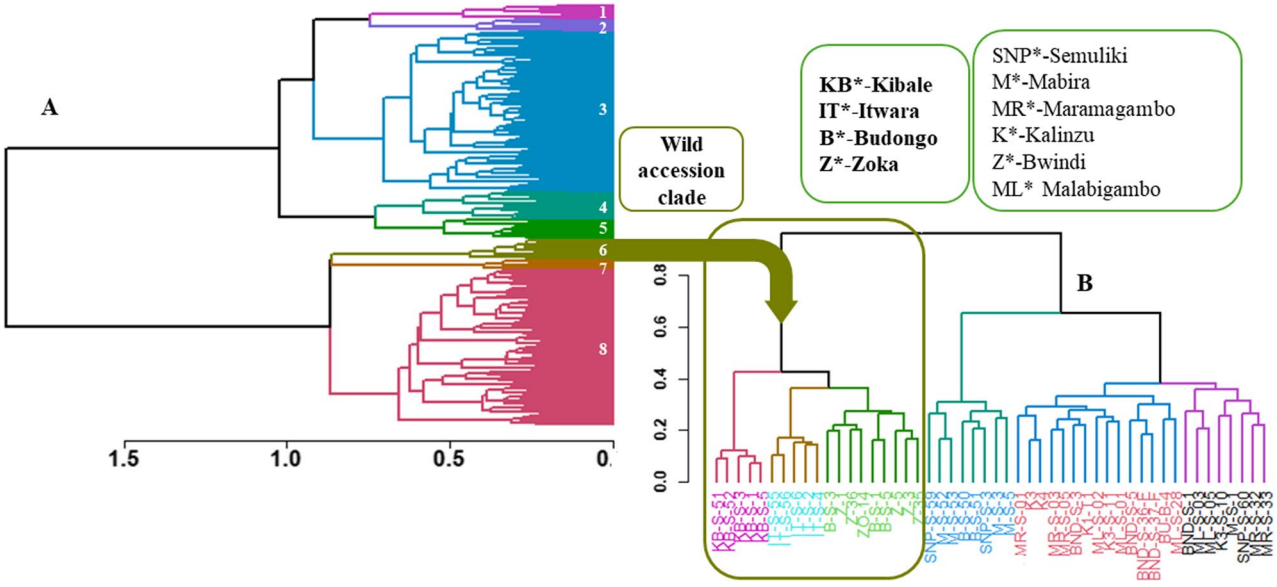

Fig. 5 Genetic relationship between the NaCORI collection and representative Ugandan wild *C. canephora*. (**A**) dendrogram of both wild and cultivated *C. canephora* accessions showing a unique cluster (cluster 6) with only wild accessions; (**B**) dendrogram of wild representative accessions

O (Uganda). Its close genetic affinity with groups E, B and R (from DRC), reflects the well-documented relationships among *C. canephora* genetic groups from this region [23, 29]. Collectively, these patterns confirm that the cultivated accessions conserved at NaCORI is predominantly of Ugandan origin, with no evidence of introduction or introgression from Guinean lineages (Fig. 2A). In this context, the Java-derived accessions, while historically associated with early introductions from Yangambi (DRC), are broadly intermixed with Ugandan cultivated genotypes and do not show a strong genetic proximity with the Congolese reference group (Fig. 2B). This pattern is consistent with historical records indicating that Robusta germplasm introduced into Java originated from multiple African sources, including Uganda [26, 67]. Accordingly, the accessions subsequently introduced into Uganda may reflect earlier bidirectional exchanges among African and Asian populations rather than derivation from a single, well-defined Congolese origin. On the other hand, the obligate outcrossing reproductive system of *C. canephora*, combined with seed-based propagation of the collection, likely promoted genetic admixture and recombination, thereby progressively eroding the original genetic identity of accessions labelled as "Java" which may in fact represent admixed descendants, as previously suggested by [23]. Collectively, these patterns indicate that the cultivated accessions conserved at NaCORI is predominantly of Ugandan origin, with no evidence for substantial introduction or introgression from non-Ugandan lineages, aside from limited contributions from DRC or Java-derived material that is largely admixed.

Collections in other African countries where *C. canephora* is native also predominantly conserve local diversity [27]. Some collections also exhibit a broader global diversity incorporating both Guinean and Congolese genetic sources but usually represented by few individuals per genetic group, for example, the Ivorian one [19, 26]. In non-African countries where coffee is not indigenous, patterns differ depending on the introduction history [68]. In the main Robusta-producing countries, Vietnam and Brazil, collection material is largely derived from Congolese group E (DRC) lineages in Vietnam [22] at the WASI (Western Highlands Agriculture and Forestry Science Institute) center and Congolese - Conilon group A lineages in Brazil at the IAC (Agronomic Institute of Campinas) collection [69], although representatives of other genetic groups are also present [22]. Assessing the complementarity of these collections will therefore be important for the long-term conservation of *C. canephora* genetic resources worldwide [39].

Geographic provenance and genetic structure of Ugandan cultivated *C. canephora* conserved at NaCORI

Cultivated germplasm showed a continuous genetic distribution, with no clear structuring by agroecological zone of origin (Fig. 3A) and displayed higher genetic diversity within regions than among them. Individuals from different agroecological were widely interspersed in multivariate space and hierarchical clustering (Fig. 3C), and only limited genetic divergence was observed, involving a subset of Southwestern accessions (Fig. 3B, C). These patterns align with previous findings in Ugandan Robusta, where the few (n = 52) cultivated accessions

studied exhibited low-to-moderate genetic structure [21, 23]. However, our study is based on a more comprehensive characterization of the national collection, encompassing a larger and more representative sample of source collection sites and agroecological zones, thereby providing a more robust and expanded perspective. This homogeneity suggests shared ancestry, recurrent germplasm exchange and seed propagation, likely facilitated by centralized seed systems, farmer-to-farmer diffusion, and institutional breeding programs. It should be noted that seed-based propagation of the collection, even over a few generations, likely intensified genetic admixture, further obscuring the genetic signatures of geographic origin. This tells the story of Uganda's coffee distribution with the central region including L. Victoria H and L. Victoria as the oldest and main distribution hub [29]. *C. canephora* has a long cultural history, particularly in regions surrounding Lake Victoria, with the Kalangala Islands representing one of the earliest centres of Robusta cultivation [28]. The use of coffee in rituals, together with its spread through human migration and intertribal trade, likely enhanced regional gene flow, contributing to the weak genetic structuring observed among cultivated germplasm [21, 28]. A similar pattern has been identified in avocado (*Persea americana*) in Tanzania [70] or *Camellia sinensis* in southwestern China [71], where cultural and social practices associated with long cultivation histories have promoted geneflow while reducing population structure [70, 71].

However, we detected a slight differentiation of a subset of accessions originating from the southwestern agroecological zone relative to the rest of the collection (Fig. 3B). These accessions form a distinct cluster (Cluster 1) in the dendrogram (Fig. 3C), whereas the remaining accessions from this zone are distributed across the other clusters, indicating an absence of clear geographic structuring. That suggests that while cultivated populations are broadly dispersed in the country and genetically similar, Southwestern accessions retain some distinct geographic genetic signatures. This distinctness (see F_{ST} values in S2 Table) may reflect limited breeding interventions and underutilized gene pools that may contain ancestral alleles. It could also result from the relative geographic isolation from traditional coffee trade routes, the presence of relatively isolated communities and ethnic groups, and the differentiation of local coffee varieties that were either poorly suited or less attractive for cultivation in other regions. These results underscore the importance of the comprehensive, country-wide sampling that was applied for the effective development of this germplasm collection. To further refine this interpretation and more robustly reconstruct such exchange events, additional sampling of extant local and traditional diversity from villages across the country could provide valuable insights.

Characterization of genetic diversity and geographic origins of the optimized core collections

Following optimization for diversity representativeness, two core collections comprising 109 and 328 accessions were identified. Core3 (109 accessions) represented the most efficient subset, maximizing allele coverage with minimal redundancy and retaining high genetic diversity (CV = 0.981; H_E = 0.294). To ensure broader geographic and genetic representativeness and increased statistical power, the larger core (Core12; 328 accessions) was designated as the GWAS population for downstream association analyses. During the optimization process, allelic coverage (CV) stabilized from Core3 onward, indicating that most alleles present in the full dataset were already captured by the third core subset and that subsequent additions did not improve representativeness in terms of allele presence. The observed reductions in the Shannon diversity index (SH) and expected heterozygosity (H_E) beyond Core3 suggest that additional accessions introduced genetic redundancy, contributing less novel diversity and skewing allele frequencies toward common variants. This pattern aligns with the principle of diminishing marginal diversity, where the initial accessions selected into a core are typically the most genetically distinct, and later additions tend to be more like those already included. Such trends were predicted by [72], who, in developing Core Hunter II, demonstrated that multi-objective optimization could identify small core sets that retain high diversity. Similarly [73], reported that in rice, core sizes of 20–30% preserved up to 95% of microsatellite alleles but were accompanied by declines in diversity indices beyond this threshold, a pattern also observed by [74] in barley.

Nonetheless, the largest core collection, Core12 (328 accessions), provided a more balanced and comprehensive representation of Uganda's genetic diversity than Core3, both in terms of geographic (district) coverage and genetic cluster composition. In addition, this core collection, by virtue of its larger size, is expected to be more effective at capturing rare allelic variants or extreme genotypes; however, this expectation cannot be robustly assessed given the limited number of SNP loci (236 analysed in this study). These results align with broader observations in the literature on core-collection. Verleysen characterised a robusta coffee field-gene bank [27]. That study emphasised that while a core can capture most allelic diversity, some specific alleles remain unrepresented. These findings reinforce the notion that core collection strategies must balance efficiency (fewer accessions) and comprehensiveness (inclusion of rare alleles). Especially for perennial crops such as

coffee, where breeding resources and field-space are constrained, core collections offer a pragmatic approach, but one must remain vigilant about losing rare alleles which might be important for future climate-resilience or disease-resistance traits as well as flavour and cup quality. In addition to the limitation that the use of hundreds of SNPs provides a restricted representation of global genetic diversity, other criteria could be incorporated into the characterization of the collections. In particular, the inclusion of phenotypic data [75] could offer complementary information and improve the overall assessment of diversity [76]. Nevertheless, this collection has strong potential as a GWAS population for several reasons. First, its adequate sample size provides sufficient statistical power to detect loci with small effects [77]. Second, it retains substantial genetic diversity, capturing broad allelic variation across multiple agroecological origins, thereby increasing the likelihood of identifying polymorphisms relevant to target traits [78]. Third, the panel is expected to exhibit moderate linkage disequilibrium (LD) [79] as suggested by the high level of admixture and limited population structure, reducing the risk of spurious associations [80]. Moreover, the ease of clonal propagation in coffee enables replicated trials across multiple agroecological environments, further enhancing the robustness of GWAS analyses.

Wild germplasm as genetic reservoirs

To assess the potential of wild *C. canephora* resources as additional germplasm for crop improvement, we evaluated representative accessions from 10 forest populations. While several wild accessions were interspersed throughout the dendrogram alongside cultivated materials, others, originating from the Kibale, Itawara, Budongo, and Zoka forests, formed a well-defined and distinct cluster, indicating genetic divergence from the cultivated gene pool (Fig. 5). This divergence indicates that part of the genetic diversity present in wild populations is not currently captured within the *ex situ* collection, highlighting their value as complementary genetic resources. In contrast, wild accessions distributed across multiple clusters together with cultivated materials likely reflect multiple domestication origins, and/or past hybridization events between wild and cultivated gene pools. These findings are consistent with studies of *C. canephora* population structure in Uganda, where four genetically distinct wild populations from northwestern forests (Zoka, Budongo, Itwara, and Kibale) were clearly separated from a fifth, more admixed Southern-Central (SC) group, which included both cultivated and wild accessions [23]. Similarly, in the DRC, wild and cultivated populations have been shown to exhibit high genetic differentiation and possess private alleles, reinforcing the idea that domestication did not arise from a single wild source but likely involved multiple wild gene pools [67]. This pattern is typical in perennial tropical crops that have undergone long-term in situ management, allowing for both divergence and convergence of gene pools [81].

Moreover, the genetic proximity of some wild accessions to cultivated accessions suggests that introgression efforts involving these materials may be feasible with minimal genetic barriers, reducing the risk of incompatibility. Deploying genetically unique germplasm given the increasing unpredictability of climate regimes, is essential for improving sustainability and crop resilience [82, 83]. Furthermore, enriched cluster-based grouping of accessions offers breeders a systematic framework to explore both intra and inter-cluster variation. This approach could enable the maximization of heterosis and facilitates the introgression of novel alleles into cultivated populations [26]. By leveraging genetic clustering, breeding programs can strategically combine diverse germplasm to generate robust varieties capable of thriving under heterogeneous agroecological conditions.

Conclusion

Using a targeted KASPar SNP panel, we characterized the genetic origin, structure, and diversity of *C. canephora* accessions representing the national collection ($n = 909$) and compared them with reference accessions from wild Ugandan populations and African reference groups. By positioning Ugandan accessions alongside reference groups from across the native range, we clarified the distinctiveness of Ugandan germplasm while also highlighting the shared ancestry of materials historically introduced from Java. Patterns of structure across landraces, farmer selections, elite breeding lines, and released varieties further revealed how collection sources and past movements have shaped the diversity now maintained *ex situ*. Kinship and redundancy analyses identified both unique and closely related accessions, which enabled the development of representative core subsets that efficiently capture the breadth of diversity while reducing duplication, an essential step for germplasm management, trait evaluation, and the design of GWAS populations. Furthermore, the inclusion of wild Ugandan populations demonstrated that natural forests remain a critical reservoir of diversity and elevated heterozygosity not fully represented in the *ex situ* gene pool.

Although this study clarified the genetic composition of the NaCORI germplasm collection and established representative core subsets, further work is needed to strengthen inference and broaden utility. Future studies integrating the characterization of the identified genetic clusters and core collections with morphological and adaptation to biotic and abiotic stresses are recommended. Moreover, this study underscores the need to integrate formal breeding programs with local seed

systems for effective germplasm utilization and highlights the value of deliberate stratification and targeted sampling from wild and under-represented cultivated populations to complement the established core subsets. Overall, these results show that the NaCORI collection provides a strong and representative foundation for Robusta improvement in Uganda, while strategic inclusion of additional wild and under-represented germplasm offers clear opportunities to further expand its genetic base.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12870-026-09023-6.

Supplementary Material 1.

Supplementary Material 2.

Supplementary Material 3.

Acknowledgements
We extend our sincere thanks to the dedicated data collectors who contributed to assembling the samples, and to the technicians from National Coffee Research Institute (NaCORI) who managed and maintained the field gene bank, ensuring the success of this work. Their commitment and expertise were essential to both the research and conservation efforts described in this study.

Authors' contributions
Conceptualization: ** Mildred Julian Nakanwagi, Pascal Musoli, Pierre Maraccini, Peter Wasswa, Catherine Kiwuka, Jeninah Karungi, Godfrey Arinaitwe, Godfrey Sseremba, Valerie Poncet, Boris DelahaieData curation & Analysis: ** Mildred Julian Nakanwagi, Mathieu Gonin, Remi Auguste, Pascal Musoli, Godfrey Ssermeba, Ali Milton, Catherine Kiwuka, Valerie Poncet, Boris Delahaie, Claude Patrick Millet, Tram Bao Vi**Investigation: ** Mildred Julian Nakanwagi, Remi Auguste, Pascal Musoli, Peter Wasswa, Catherine Kiwuka, Godfrey Sseremba, Valerie Poncet, Boris Delahaie**Project administration: ** Mildred Julian Nakanwagi, Pascal Musoli, Pierre Maraccini, Peter Wasswa, Jeninah Karungi, Mildred Ochwo Ssemakula, Godfrey Arinaitwe, Godfrey Sseremba, Valerie Poncet**Visualization: ** Mildred Julian Nakanwagi, Mathieu Gonin, Peter Wasswa, Valerie Poncet, Boris Delahaie**Writing and reviewing** : Mildred Julian Nakanwagi, Mathieu Gonin, Pascal Musoli, Pierre Maraccini, Remi Auguste, Tram Bao Vi, Ali Milton, Peter Wasswa, Catherine Kiwuka, Godfrey Arinaitwe, Jeninah Karungi, Isaac Dramadri, Mildred Ochwo Ssemakula, Claude Patrick Millet, Godfrey Sseremba, Valerie Poncet, Boris Delahaie.

Funding
This work was funded at NaCORI by the European Union (EU) for the project "Robusta coffee agroforestry to adapt and mitigate climate change in Uganda" (GCCA+-Global Climate Change Alliance & DESIRA, Project/Grant No. FOOD/2021/427–759. MJN received additional funding from the ARTS program (IRD) and the "Action incitatives" program from the Centre de Coopération Internationale en Recherche Agronomique pour le Développement (CIRAD).

Data availability
All data supporting the findings of this study are available within the paper and its Supplementary Information. KASPar marker sequences are provided in Supplementary data (Excel sheet) attached, along with original reference describing the SNPs used in this study.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Author details
[1]Department of Crop Science and Horticulture, Makerere University, Kampala, Uganda
[2]National Coffee Research Institute, National Agricultural Research Organisation, Kituza, Mukono, Uganda
[3]CIRAD, UMR DIADE, Montpellier F-34398, France
[4]UMR DIADE, Univ Montpellier, IRD, CIRAD, Montpellier, France
[5]Plant Genetic Resource Center, National Agricultural Research Organisation, Entebbe, Uganda
[6]Faculty of Agricultural Sciences, Uganda Christian University, Mukono, Uganda
[7]Univ Paris-Est Créteil, Institute of Ecology and Environmental Science, IEES, Créteil, France

Received: 13 January 2026 / Accepted: 18 May 2026
Published online: 25 May 2026

References

1. Raza A, Razzaq A, Mehmood SS, Zou X, Zhang X, Lv Y, Xu J. Impact of Climate Change on Crops Adaptation and Strategies to Tackle Its Outcome: A Review. Plants. 2019;8:34. https://doi.org/10.3390/plants8020034.
2. Weiskopf SR, Rubenstein MA, Crozier LG, Gaichas S, Griffis R, Halofsky JE, Hyde KJW, Morelli TL, Morisette JT, Muñoz RC, et al. Climate Change Effects on Biodiversity, Ecosystems, Ecosystem Services, and Natural Resource Management in the United States. Sci Total Environment. 2020;733:137782. https://doi.org/10.1016/j.scitotenv.2020.137782.
3. Chapman EA, Thomsen HC, Tulloch S, Correia PMP, Luo G, Najafi J, DeHaan LR, Crews TE, Olsson L, Lundquist P-O, et al. Perennials as Future Grain Crops: Opportunities and Challenges. Front Plant Sci. 2022;13:898769. https://doi.org/10.3389/fpls.2022.898769.
4. Jungers J, Runck B, Ewing PM, Maaz T, Carlson C, Neyhart J, Fumia N, Bajgain P, Subedi S, Sharma V, et al. Adapting Perennial Grain and Oilseed Crops for Climate Resiliency. Crop Sci. 2023;63:1701–21. https://doi.org/10.1002/csc2.20972.
5. Jarvis A, Upadhyaya H, Gowda CLL, Aggarwal PK, Fujisaka S, Anderson B. Climate change and its effect on conservation and use of plant genetic resources for food and agriculture and associated biodiversity for food security. Intern ational Centre for Tropical Agriculture (CIAT); Bioversity International; ICRISAT; Indian Agricultural Research Institute; American University; 2010. https://hdl.handle.net/10568/33384.
6. Khoury CK, Brush S, Costich DE, Curry HA, de Haan S, Engels JMM, Guarino L, Hoban S, Mercer KL, Miller AJ, et al. Crop Genetic Erosion: Understanding and Responding to Loss of Crop Diversity. New Phytol. 2022;233:84–118. https://doi.org/10.1111/nph.17733.
7. Salgotra RK, Chauhan BS. Genetic Diversity, Conservation, and Utilization of Plant Genetic Resources. Genes (Basel). 2023;14:174. https://doi.org/10.3390/genes14010174.
8. Dulloo ME, Solano W, Dessauw D, Astorga C, Guarino L. A Methodological Approach for Prioritization and Rationalization of Field Gene bank Accessions of Coffee Genetic Resources: A Case Study of CATIE International Coffee Collection, Costa Rica. Front Sustain Food Syst. 2021;5:777415. https://doi.org/10.3389/fsufs.2021.777415.
9. Maxted N, Adam-Blondon A-F, Aguilar CH, Barata AM, Bartha B, Bocci R, Paola DD, Fitzgerald HS, Fresta LJ, Fusani P, et al. A significantly enhanced role for plant genetic resource centres in linking *in situ* and *ex situ* conservation to aid user germplasm access. Genetic Resour. 2024;203–23. https://doi.org/10.46265/genresj.UNVV5571.
10. Cortés AJ, López-Hernández F. Harnessing Crop Wild Diversity for Climate Change Adaptation. Genes. 2021;12:783. https://doi.org/10.3390/genes12050783.
11. Hagenblad J, Aloisi K, Marum P, Öhlund L, Solberg SØ, Asdal Å, Palmé A. Limited genetic changes observed during in situ and ex situ conservation in nordic populations of red clover (*Trifolium pratense*). Front Plant Sci. 2023;14. https://doi.org/10.3389/fpls.2023.1233838.

12. Engels JMM, Thormann I. Main Challenges and Actions Needed to Improve Conservation and Sustainable Use of Our Crop Wild Relatives. Plants (Basel). 2020;9:968. https://doi.org/10.3390/plants9080968.
13. Hyten DL, Song Q, Zhu Y, Choi I-Y, Nelson RL, Costa JM, Specht JE, Shoemaker RC, Cregan PB. Impacts of Genetic Bottlenecks on Soybean Genome Diversity. Proc Natl Acad Sci. 2006;103:16666–71. https://doi.org/10.1073/pnas.0604379103.
14. Smýkal P, Nelson M, Berger J, Von Wettberg E. The Impact of Genetic Changes during Crop Domestication. Agronomy. 2018;8:119. https://doi.org/10.3390/agronomy8070119.
15. Davis AP, Govaerts R, Bridson DM, Stoffelen P. An annotated taxonomic conspectus of the Genus *Coffea* (*Rubiaceae*). Bot J Linn Soc. 2006;152:465–512, https://doi.org/10.1111/j.1095-8339.2006.00584.x.
16. Tournebize R, Chu G, Moorjani P. Reconstructing the History of Founder Events Using Genome-Wide Patterns of Allele Sharing across Individuals. PLoS Genet. 2022;18:e1010243. https://doi.org/10.1371/journal.pgen.1010243.
17. Depecker J. Genetic diversity and structure in wild robusta coffee (*Coffea canephora* A. Froehner) populations in Yangambi (DR Congo) and their relation to forest disturbance. Heredity 2023, https://doi.org/10.1038/s41437-022-00588-0.
18. Berthaud J. Les Ressources Genetiques Pour l'amelioration Des Cafeires Africains Diploides; ORSTOM: Paris. 1986;188;ISBN 2-7099-0811-5.
19. Gomez C, Dussert S, Hamon P, Hamon S, de Kochko A, Poncet V. Current Genetic Differentiation of *Coffea canephora* Pierre Ex A. Froehn in the Guineo-Congolian African Zone: Cumulative Impact of Ancient Climatic Changes and Recent Human Activities. BMC Evol Biol. 2009;9:167.
20. Mérot-L'Anthoëne V, Tournebize R, Darracq O, Rattina V, Lepelley M, Bellanger L, Tranchant-Dubreuil C, Coulée M, Pégard M, Metairon S, et al. Development and Evaluation of a Genome-wide Coffee 8.5K SNP Array and Its Application for High-density Genetic Mapping and for Investigating the Origin of *Coffea arabica* L. Plant Biotechnol J. 2019;17:1418–30. https://doi.org/10.1111/pbi.13066.
21. Musoli P, Cubry P, Aluka P, Billot C, Dufour M, De Bellis F, Pot D, Bieysse D, Charrier A, Leroy T. Genetic Differentiation of Wild and Cultivated Populations: Diversity of *Coffea Canephora* Pierre in Uganda. Genome. 2009;52:634–46. https://doi.org/10.1139/G09-037.
22. Vi T, Le TN, Cubry P, Phan VH, Dinh TTO, Tran TBN, Nguyen VT, Millet CP, Kambale J-L, Karine GKM, et al. Out of Africa: The Genomic Footprints of Vietnamese Robusta coffee. PLoS ONE. 2025;20:e0324988. https://doi.org/10.1371/journal.pone.0324988.
23. Kiwuka C, Goudsmit E, Tournebize R, Aquino SO, Douma JC, Bellanger L, Crouzillat D, Stoffelen P, Sumirat U, Legnaté H, et al. Genetic Diversity of Native and Cultivated Ugandan Robusta coffee (*Coffea canephora* Pierre Ex A. Froehner): Climate Influences, Breeding Potential and Diversity Conservation. PLoS ONE. 2021;16:e0245965. https://doi.org/10.1371/journal.pone.0245965.
24. Thomas AS. The wild coffees of Uganda. Emp J Exp Agric. 1944:1–12.
25. Cramer PJS. A review of literature of coffee research in Indonesia; SIC Editorial, Inter American Institute of Agricultural Sciences: Turrialba, Costa Rica. 1957;262.
26. Montagnon C, Leroy T, Eskes A. Varietal improvement of *Coffea canephora*. 1: criteria and breeding methods. Recherche: Plantations; Développement. 1998. http://agritrop.cirad.fr/390311/.
27. Verleysen L, Bollen R, Kambale J-L, Ebele T, Katshela BN, Depecker J, Poncet V, Assumani D-M, Vandelook F, Stoffelen P et al. Characterization of the genetic composition and establishment of a core collection for the INERA Robusta coffee (*Coffea canephora*) field gene bank from the Democratic Republic of Congo. Front Sustainable Food Syst. 2023;7. https://doi.org/10.3389/fsufs.2023.1239442.
28. Thomas AS. Types of Robusta coffee and their Selection in Uganda. East Afr Agricultural J. 1935;1:193–7. https://doi.org/10.1080/03670074.1935.11663646.
29. Musoli P. Recherche de Sources de Résistance à La Trachéomycose Du Caféier *Coffea Canephora* Pierre. Due à Fusarium xylarioides Steyaert En Ouganda. 2007. https://institut-agro.docressources.fr/index.php?lvl=notice_display&id=40386.
30. Kibirige-Ssebunya I, Nabasirye M, Matovu J, Musoli PA. Comparison Among Various Robusta Coffee (*Coffea canephora* Pierre) Clonal Materials and Their Seedling Progenies at Different Levels of Nitrogen. Uganda J Agricultural Sci. 1993;1:5–12.
31. Kassa A, Akachukwu D, Kadege EL, Konan-Kakou JA, Teklay A, Ojimelukwe PC. Climate Change Impacts and Adaptations in Eastern and Western Africa Crop Production Systems: Lessons and Transferable Models among Smallholder Farmers. Clim Serv. 2026;42:100644. https://doi.org/10.1016/j.cliser.2026.100644.
32. Jaramillo J, Muchugu E, Vega FE, Davis A, Borgemeister C, Chabi-Olaye A. Some Like It Hot: The Influence and Implications of Climate Change on Coffee Berry Borer (*Hypothenemus hampei*) and Coffee Production in East Africa. PLoS ONE. 2011;6:e24528. https://doi.org/10.1371/journal.pone.0024528.
33. Engels JMM, Ebert AW. A Critical Review of the Current Global Ex Situ Conservation System for Plant Agrobiodiversity. I. History of the Development of the Global System in the Context of the Political/Legal Framework and Its Major Conservation Components. Plants. 2021;10:1557. https://doi.org/10.3390/plants10081557.
34. Aluka P, Davrieux F, Ngugi K, Omaria R, Gerard F, Leroy T, Dufour M, Fabrice P. The diversity of green bean biochemical compounds in Robusta coffee (*Coffea canephora* Pierre Ex A. Froehner) as evaluated by near infrared spectroscopy. AJEA. 2016;12:1–13. https://doi.org/10.9734/AJEA/2016/23004.
35. Anagbogu CF, Bhattacharjee R, Ilori C, Tongyoo P, Dada KE, Muyiwa AA, Gepts P, Beckles DM. Genetic Diversity and Re-Classification of Coffee (Coffea canephora Pierre Ex A. Froehner) from South Western Nigeria through Genotyping-by-Sequencing-Single Nucleotide Polymorphism Analysis. Genet Resour Crop Evol. 2019;66:685–96. https://doi.org/10.1007/s10722-019-00744-2.
36. Asefa Bikila B. SNPs Based molecular diversity of *Coffea canephora*. JMEN. 2017;5. https://doi.org/10.15406/jmen.2017.05.00136.
37. Garavito A, Montagnon C, Guyot R, Bertrand B. Identification by the DArTseq Method of the Genetic Origin of the *Coffea canephora* Cultivated in Vietnam and Mexico. BMC Plant Biol. 2016;16:242. https://doi.org/10.1186/s12870-016-0933-y.
38. Huang CQ, Long T, Bai CJ, Wang WQ, Tang J, Liu GD. Establishment of a Core Collection of *Cynodon* Based on Morphological Data. Trop grassl -Forrajes trop. 2020;8:203–13. https://doi.org/10.17138/tgft(8)203-213.
39. Vi T, Marraccini P, Kochko A. Cubry, P. Khong NG. Poncet V. Sequencing-based molecular markers for wild and cultivated coffee diversity exploration and crop improvement. In Coffee Science; CRC Press. 2022. ISBN 978-1-003-04313-3. https://doi.org/10.1201/9781003043133.
40. Akpertey A, Padi FK, Meinhardt L, Zhang D. Effectiveness of single nucleotide polymorphism markers in genotyping germplasm collections of *Coffea canephora* Using KASP Assay. Front. Plant Sci. 2021;11. https://doi.org/10.3389/fpls.2020.612593.
41. Millet CP, Allinne C, Vi T, Marraccini P, Verleysen L, Couderc M, Ruttink T, Zhang D, Sanchéz WS, Tranchant-Dubreuil C, et al. Haitian Coffee Agroforestry Systems Harbor Complex Arabica Variety Mixtures and Under-Recognized Genetic Diversity. PLoS ONE. 2024;19:e0299493. https://doi.org/10.1371/journal.pone.0299493.
42. Upadhyaya HD, Gowda CLL, Reddy KN, Singh S. Augmenting the Pearl Millet Core Collection for Enhancing Germplasm Utilization in Crop Improvement. Crop Sci. 2009;49:573–80. https://doi.org/10.2135/cropsci2008.06.0378.
43. Brown AHD, Core Collections. A practical approach to genetic resources management. Genome. 1989;31:818–24.
44. Coşkun ÖF, Molecular Characterization. Population Structure Analysis, and Association Mapping of Turkish Parsley Genotypes Using iPBS Markers. Horticulturae. 2023;9:336. https://doi.org/10.3390/horticulturae9030336.
45. Leroy T, De Bellis F, Legnate H, Musoli P, Kalonji A, Loor Solórzano RG, Cubry P. Developing Core Collections to Optimize the Management and the Exploitation of Diversity of the Coffee *Coffea canephora*. Genetica. 2014;142:185–99. https://doi.org/10.1007/s10709-014-9766-5.
46. Su T, Li P, Yang J, Sui G, Yu Y, Zhang D, Zhao X, Wang W, Wen C, Yu S, et al. Development of Cost-Effective Single Nucleotide Polymorphism Marker Assays for Genetic Diversity Analysis in *Brassica rapa*. Mol Breed. 2018;38:42. https://doi.org/10.1007/s11032-018-0795-0.
47. Kebede G, Mudereri BT, Mutanga O, Landmann T, Odindi J, Motisi N, Pinard F, Tonnang HEZ, Abdel-Rahman EM. Mapping Robusta coffee (*Coffea canephora*) Cropping Systems in Uganda: A Two-Step Pixel and Sub-Pixel Based Approach with Sentinel-2 Data. PLoS ONE. 2026;21:e0338803. https://doi.org/10.1371/journal.pone.0338803.
48. Basalirwa CPK. Delineation of Uganda into Climatological Rainfall Zones Using the Method of Principal Component Analysis. Int J Climatol. 1995;15:1161–77. https://doi.org/10.1002/joc.3370151008.
49. Wortmann: Uganda's Agroecological Zones: A Guide… Google Scholar Available online: https://scholar.google.fr/scholar?cluster=5851947971570318403&hl=en&as_sdt=0,5. Accessed on 10 November 2025.

50. Bernard B, Anthony E, Patrick O, Fridah M. Dynamics of land use/cover trends in Kanungu District, South Western Uganda. J Appl Sci Environ Manage. 2010;14. https://doi.org/10.4314/jasem.v14i4.63260.
51. Plechakova O, Tranchant-Dubreuil C, Benedet F, Couderc M, Tinaut A, Viader V, De Block P, Hamon P, Campa C, de Kochko A, et al. MoccaDB - an Integrative Database for Functional, Comparative and Diversity Studies in the Rubiaceae Family. BMC Plant Biol. 2009;9:123.
52. Gezan SA, de Oliveira AA, Galli G, Murray D. User's manual for ASRgenomics v. 1.1.0 An R package with complementary genomic functions. 2022 https://asreml.kb.vsni.co.uk/wp-content/uploads/sites/3/ASRgenomics_Manual.pdf.
53. R Core Team. (2023) R A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna. - References - Scientific Research Publishing Available online: https://www.scirp.org/reference/referencespapers?referenceid=3582659. Accessed on 14 November 2025.
54. Jombart T, Adegenet. A R Package for the Multivariate Analysis of Genetic Markers. Bioinformatics. 2008;24:1403–5. https://doi.org/10.1093/bioinformatics/btn129.
55. Goudet J, Hierfstat. a Package for r to Compute and Test Hierarchical F-Statistics. Mol Ecol Notes. 2005;5:184–6. https://doi.org/10.1111/j.1471-8286.2004.00828.x.
56. Zheng X, Levine D, Shen J, Gogarten SM, Laurie C, Weir BS. A High-Performance Computing Toolset for Relatedness and Principal Component Analysis of SNP Data. Bioinformatics. 2012;28:3326–8. https://doi.org/10.1093/bioinformatics/bts606.
57. Frichot E, Mathieu F, Trouillon T, Bouchard G, François O. Fast and Efficient Estimation of Individual Ancestry Coefficients. Genetics. 2014;196:973–83. https://doi.org/10.1534/genetics.113.160572.
58. Wickham H, Averick M, Bryan J, Chang W, McGowan LD, François R, Grolemund G, Hayes A, Henry L, Hester J, et al. Welcome to the Tidyverse. J Open Source Softw. 2019;4:1686. https://doi.org/10.21105/joss.01686.
59. Müllner D, Fastcluster. Fast hierarchical, agglomerative clustering routines for *R* and *Python*. J Stat Soft. 2013;53. https://doi.org/10.18637/jss.v053.i09.
60. Galili T, Dendextend. An R Package for Visualizing, Adjusting and Comparing Trees of Hierarchical Clustering. Bioinformatics. 2015;31:3718–20. https://doi.org/10.1093/bioinformatics/btv428.
61. Weir BS, Cockerham CC. Estimating F-Statistics for the Analysis of Population Structure. Evolution. 1984;38:1358–70. https://doi.org/10.1111/j.1558-5646.1984.tb05657.x.
62. De Beukelaer H, Davenport GF, Fack VC. Hunter 3: Flexible Core Subset Selection. BMC Bioinformatics. 2018;19:203. https://doi.org/10.1186/s12859-018-2209-z.
63. Odong TL. Quantitative methods for sampling of germplasm collections: getting the best out of molecular markers when creating core collections. Wageningen University. 2012. https://research.wur.nl/en/publications/quantitative-methods-for-sampling-of-germplasm-collections-gettin/.
64. Gu R, Fan S, Wei S, Li J, Zheng S, Liu G. Developments on Core Collections of Plant Genetic Resources. Do We Know Enough? Forests. 2023;14:926. https://doi.org/10.3390/f14050926.
65. Franco J, Crossa J, Warburton ML, Taba S. Sampling Strategies for Conserving Maize Diversity When Forming Core Subsets Using Genetic Markers. Crop Sci. 2006;46:854–64. https://doi.org/10.2135/cropsci2005.07-0201.
66. Thachuk C, Crossa J, Franco J, Dreisigacker S, Warburton M, Davenport GF. Core Hunter: An Algorithm for Sampling Genetic Resources Based on Multiple Genetic Measures. BMC Bioinformatics. 2009;10:243. https://doi.org/10.1186/1471-2105-10-243.
67. Vanden Abeele S, Janssens SB, Asimonyio Anio J, Bawin Y, Depecker J, Kambale B, Mwanga Mwanga I, Ebele T, Ntore S, Stoffelen P, et al. Genetic Diversity of Wild and Cultivated *Coffea canephora* in Northeastern DR Congo and the Implications for Conservation. Am J Bot. 2021;108:2425–34. https://doi.org/10.1002/ajb2.1769.
68. Loor Solórzano RG, De Bellis F, Leroy T, Plaza L, Guerrero H, Subia C, Calderón D, Fernández F, Garzón I, Lopez D, et al. Revealing the diversity of introduced *Coffea canephora* germplasm in Ecuador: towards a national strategy to improve Robusta. Sci World J. 2017;1–12. https://doi.org/10.1155/2017/1248954.
69. Ferrão MAG, Ferrão RG, Fonseca AFA, Filho AbraãoCV, Volpi PS, Origin. Geographical dispersion, taxonomy and genetic diversity of *Coffea canephora*. In Conilon coffee; Incaper. 2019. ISBN 978-85-89274-32-6. http://biblioteca.incaper.es.gov.br/digital/handle/123456789/3540.
70. Juma I, Geleta M, Nyomora A, Saripella GV, Hovmalm HP, Carlsson AS, Fatih M, Ortiz R. Genetic Diversity of Avocado from the Southern Highlands of Tanzania as Revealed by Microsatellite Markers. Hereditas. 2020;157:40. https://doi.org/10.1186/s41065-020-00150-0.
71. Xin T, Huang W, De Riek J, Zhang S, Ahmed S, Van Huylenbroeck J, Long C. Genetic Diversity, Population Structure, and Traditional Culture of *Camellia reticulata*. Ecol Evol. 2017;7:8915–26. https://doi.org/10.1002/ece3.3340.
72. Beukelaer HD, Smýkal P, Davenport GF, Fack V, Core Hunter II. Fast Core Subset Selection Based on Multiple Genetic Diversity Measures Using Mixed Replica Search. BMC Bioinformatics. 2012;13:312. https://doi.org/10.1186/1471-2105-13-312.
73. Zhang P, Li J, Li X, Liu X, Zhao X, Lu Y. Population Structure and Genetic Diversity in a Rice Core Collection (*Oryza sativa* L.) Investigated with SSR Markers. PLoS ONE. 2011;6:e27565. https://doi.org/10.1371/journal.pone.0027565.
74. Kaur V, Aravind J, Manju Jacob SR, Kumari J, Panwar BS, Pal N, Rana JC, Pandey A, Kumar A. Phenotypic characterization, genetic diversity assessment in 6,778 accessions of Barley (*Hordeum vulgare* L. ssp. *Vulgare*) germplasm conserved in National Gene bank of India and Development of a Core Set. Front Plant Sci. 2022;13. https://doi.org/10.3389/fpls.2022.771920.
75. Ali Milton, Mukasa SB, Poncet V, Sseremba G, Marraccini P, et al. Intraspecific drought tolerance in Ugandan Coffea canephora for accelerated breeding selection. PLOS ONE 2026;21(5):e0349873. https://doi.org/10.1371/journal.pone.0349873.
76. Upadhyaya HD, Dwivedi SL, Gowda CLL, Singh S. Identification of Diverse Germplasm Lines for Agronomic Traits in a Chickpea (*Cicer arietinum* L.) Core Collection for Use in Crop Improvement. Field Crops Res. 2007;100:320–6. https://doi.org/10.1016/j.fcr.2006.08.008.
77. Visscher PM, Wray NR, Zhang Q, Sklar P, McCarthy MI, Brown MA, Yang J. 10 Years of GWAS discovery: biology, function, and translation. Am J Hum Genet. 2017;101:5–22, https://doi.org/10.1016/j.ajhg.2017.06.005.
78. Zhao K, Tung C-W, Eizenga GC, Wright MH, Ali ML, Price AH, Norton GJ, Islam MR, Reynolds A, Mezey J, et al. Genome-Wide Association Mapping Reveals a Rich Genetic Architecture of Complex Traits in *Oryza sativa*. Nat Commun. 2011;2:467. https://doi.org/10.1038/ncomms1467.
79. Ingvarsson PK, Street NR. Association Genetics of Complex Traits in Plants. New Phytol. 2011;189:909–22. https://doi.org/10.1111/j.1469-8137.2010.03593.x.
80. Yu J, Pressoir G, Briggs WH, Vroh Bi I, Yamasaki M, Doebley JF, McMullen MD, Gaut BS, Nielsen DM, Holland JB, et al. A Unified Mixed-Model Method for Association Mapping That Accounts for Multiple Levels of Relatedness. Nat Genet. 2006;38:203–8. https://doi.org/10.1038/ng1702.
81. Verleysen L, Depecker J, Bollen R, Asimonyio J, Hatangi Y, Kambale J-L, Mwanga Mwanga I, Ebele T, Dhed'a B, Stoffelen P, et al. Crop-to-Wild Gene Flow in Wild Coffee Species: The Case of *Coffea canephora* in the Democratic Republic of the Congo. Ann Botany. 2024;133:917–30. https://doi.org/10.1093/aob/mcae034.
82. Lwiza F, Barkley A. Climate Variation Effect on Robusta coffee (*Coffea canephora*) Yield in Uganda. Reg Environ Change. 2025;25:50. https://doi.org/10.1007/s10113-025-02370-4.
83. Mulinde C, Majaliwa JGM, Twinomuhangi R, Mfitumukiza D, Waiswa D, Tumwine F, Kato E, Asiimwe J, Nakyagaba WN, Mukasa D. Projected Climate in Coffee-Based Farming Systems: Implications for Crop Suitability in Uganda. Reg Environ Change. 2022;22:83. https://doi.org/10.1007/s10113-022-01930-2.

Publisher's note