

RESOURCE

Scaling up orphan crop research: genebank genetics highlight geographic structure in cultivated cowpea from 10 617 global accessions

Sofie Pearson^{1,*} , Adrian Hathorn¹ , Shichao Sun² , Alan Cruickshank^{3,†} , Tracey Shatte³ , Paulino Munisse⁴, Mercy Macharia Wairimu⁵ , Joann Conner⁶ , Anna Koltunow⁷ , Jean-Philippe Vielle-Calzada⁸ , Peggy Ozias-Akins⁶ , Takayoshi Ishii^{9,10} , Matteo Dell'Acqua⁵ , Sally Norton¹¹ , Yongfu Tao² , David Jordan¹ , and Emma Mace^{1,*} 

¹Queensland Alliance for Agriculture and Food Innovation (QAAFI), The University of Queensland, Hermitage Research Facility, Warwick, Queensland 4370, Australia,

²Shenzhen Branch, Guangdong Laboratory of Lingnan Modern Agriculture, Agricultural Genomics Institute at Shenzhen, Chinese Academy of Agricultural Sciences, Shenzhen 518120, China,

³Queensland Department of Primary Industries (DPI), Hermitage Research Facility, Warwick, Queensland 4370, Australia,

⁴Mozambique Institute of Agricultural Research, Maputo, Mozambique,

⁵Institute of Plant Sciences, Scuola Superiore Sant'Anna, Pisa, Italy,

⁶Department of Horticulture and Institute of Plant Breeding, Genetics and Genomics, University of Georgia, Tifton, Georgia 31793, USA,

⁷Queensland Alliance for Agriculture and Food Innovation (QAAFI), The University of Queensland, St Lucia, Brisbane, Queensland 4072, Australia,

⁸Grupo de Desarrollo Reproductivo y Apomixis, Unidad de Genómica Avanzada Laboratorio Nacional de Genómica Para la Biodiversidad, Cinvestav Irapuato, Guanajuato CP36825, México,

⁹International Platform for Dryland Research and Education (IPDRE), Tottori University, Tottori 680-0001, Japan,

¹⁰Arid Land Research Center (ALRC), Tottori University, Tottori 680-0001, Japan, and

¹¹Australian Grains Genebank, Horsham, Victoria, Australia

Received 4 November 2025; revised 13 February 2026; accepted 17 February 2026.

*For correspondence (e-mail emma.mace@uq.edu.au and sofie.pearson@uq.edu.au).

[†]Deceased.

SUMMARY

Vigna unguiculata (L.) Walp. is a dryland legume crop, providing essential food and nutritional security for millions of people across the semi-arid tropics, in Africa, Asia and Latin America. However, as a typical 'orphan crop', cowpea has long remained underrepresented in global genomic research to support crop improvement. Here, we conducted the largest genetic diversity analysis of cowpea to date, comprising 10 617 accessions sourced from seven international collections. Using genotyping-by-sequencing, we characterised the global patterns of genetic diversity, assessed redundancy within and across collections, and examined the geographic structure of the cowpea global allele pool. Our results revealed nine distinct genetic groups with clear geographic associations and fine-scale population differentiation, reflecting dispersal history, regional adaptation and the influence of modern breeding. Duplication across collections was detected, highlighting the need for improved curation and integration of germplasm resources. Landraces from sub-Saharan Africa do not fully capture the genetic diversity present in several other geographic regions, indicating the existence of abundant and untapped genetic resources worldwide. These findings not only provide insights into the genetic structure and evolutionary history of cowpea but also offer a valuable foundation for harnessing global germplasm diversity to enhance breeding potential and accelerate crop improvement.

Keywords: genetic structure, legume breeding, duplicates, geographic structure, biogeography, orphan crop, opportunity crop, germplasm, pulse, phylogenetics.

INTRODUCTION

Global food security remains a pressing issue, with an estimated 1 billion people currently underfed, a situation that is likely to intensify as the global population is predicted to surpass 9 billion by 2050 (Gu et al., 2021). Meeting this future demand will require increasing food production while meeting sustainability objectives (Tilman et al., 2011; Van Dijk et al., 2021). However, relying solely on improvements in major staple crops is unlikely to meet this target (Ray et al., 2013). Diversifying food sources is therefore imperative to ensure nutritional security (Gruber, 2017). Cultivating orphan crops (those which lack advanced scientific research and breeding) offers a promising avenue for diversification and sustainability. Orphan crops include cereals, legumes and root crops often cultivated in limited areas in Africa, Asia and Latin America, play an important cultural role and are vital for the nutrition and livelihood of many households (Ye & Fan, 2021). These crops possess desirable traits such as resilience to extreme climatic conditions and high nutritional value (Tadele, 2019) and are also referred to as 'opportunity crops' (Yang et al., 2025). Additionally, plant genetic resources including landraces and wild crop relatives conserved in international genebanks are a valuable reservoir of genetic diversity, offering access to novel allelic variation that could contribute to sustainable intensification of yields, providing resistance to diseases and abiotic stress, and increasing nutritional quality (Ye & Fan, 2021).

Genebanks play a critical role in safeguarding plant genetic resources by storing seeds from diverse species, including orphan crops. These large diverse collections promote conservation and support plant breeding and crop improvement, especially in the face of changing environmental conditions (Asdal & Guarino, 2018). This includes maintaining collections of crop seeds, including traditional and heirloom varieties that may possess traits such as pest resistance, drought tolerance and adaptability to low-input environments (Hay & Probert, 2013). Effective management of these genetic reservoirs is crucial for developing cultivars capable of addressing emerging agricultural challenges (Tanksley & Mccouch, 1997). Therefore, accurate identification and conservation of genetic diversity within collections are essential for optimising breeding programmes (Mascher et al., 2019). Previous studies assessing diversity in genebanks include major staple crops such as wheat (Schulthess et al., 2022), barley (Milner et al., 2019) and rice (Mccouch et al., 2012), but limited studies have assessed the diversity in orphan crops.

Cultivated cowpea is a globally important grain legume, cultivated on approximately 15 million hectares, producing on average 9.7 million tonnes in the past 5 years (FAO, 2025) and provides a major source of nutrition and income for hundreds of millions of households in the

semi-arid tropics, including Africa, Asia and Latin America (Boukar et al., 2019; Dixit & Satheesh Naik, 2025). Despite this importance, cowpea has historically received comparatively limited breeding research investment and genomic resource development relative to major legumes such as soybean, common bean and chickpea. This genomic underrepresentation has led to its classification as an 'orphan crop'. However, cowpea now occupies a transitional zone, where research is still underfunded relative to its significance but advances in international breeding efforts (Kim et al., 2025) and genomic resources (Liang et al., 2024; Lonardi et al., 2019) are increasingly supported. As genomic resources expand, well-characterised genebank collections remain essential for accessing the genetic diversity needed to drive further improvement in cowpea. The largest genebank collections of *Vigna unguiculata* (L.) Walp include the International Institute of Tropical Agriculture (IITA) in Nigeria, the United States Department of Agriculture (USDA) National Plant Germplasm System in the USA and the National Bureau of Plant Genetic Resources (NBPGR) in India. Cultivated cowpea are grouped under *V. unguiculata* subsp. *unguiculata*, which is subdivided into five cultivar groups (c.g.) based on seed and pod traits: (1) *Unguiculata* (cowpea or black-eyed bean, widely cultivated as a pulse or vegetable with more than 16 ovules per pod); (2) *Biflora/Cylindrica* (catjang, common in India, primarily used for forage and has short erect pods with less than 17 ovules per pod and smooth seed); (3) *Sesquipedalis* (yardlong or asparagus bean, common in Asia, grown as a vegetable with very long pods; also known as *Vigna unguiculata* subsp. *sesquipedalis* (L.) Verdc.); (4) *Textilis* (a rare form once cultivated for fibre in Africa with long floral peduncles); and (5) *Melanophthalmus* (grown mostly in the Americas, has less than 17 ovules per pod and includes blackeye pea types) (Pasquet, 2000). Cowpea seeds, leaves and green pods are all utilised and provide essential micronutrients such as zinc and iron, which are often lacking in other legumes (Abebe & Alemayehu, 2022; Jayathilake et al., 2018). Comprehensive characterisation of genebank collections can increase improvement in cultivar development, trait discovery and translate genomic advancements into practical breeding outcomes.

In sub-Saharan Africa, smallholder farmers are the main producers and consumers of cowpea (Bolarinwa et al., 2022) and often grow it as an intercrop with cereals and other crops (Ongom et al., 2023). Cowpea compensates for the nitrogen loss by cereals through sustainable nitrogen fixation in farming systems (Kussie et al., 2024). Cowpea's centre of domestication is in Africa; from there it has spread to all continents and exhibits high resilience to drought environments and poor soil conditions. However, yields are often adversely affected by cowpea's susceptibility to pests (such as aphids, weevils and pod

borers) and diseases (such as Anthracnose and fusarium wilt) (Singh & Allen, 1979). These yield challenges pose a threat to global food security, highlighting the need for continuous breeding efforts. Characterising global germplasm is critical to explore the untapped resource to meet the challenges in cowpea breeding.

Over the last few decades, significant progress has been made in characterising the genetic diversity of cowpea. Genetic studies have consistently highlighted West Africa as the primary centre of cowpea diversity, with substantial variation also found across Southern and Eastern Africa, Asia and the Americas (Fang et al., 2007; Fiscus et al., 2024; Huynh et al., 2013; Sodedji et al., 2021; Xiong et al., 2016). Multiple studies have focused on the diversity in specific geographic regions, including Western Africa (Egbadzor et al., 2014, Gbedevi et al., 2021), Southern Africa (Chipeta et al., 2025; Macharia et al., 2025; Nkhoma et al., 2020), Eastern Africa (Desalegne et al., 2016; Ketema et al., 2020), the Iberian Peninsula (Carvalho et al., 2017), China (Chen et al., 2017) and Korea (Seo et al., 2020). Alongside these diversity studies, cowpea breeding programmes have been established across West, Central, Eastern and Southern Africa, as well as in Brazil, the USA and parts of Asia (Singh et al., 1997). These programmes typically draw upon regionally adapted landraces and elite lines to target key agronomic constraints such as drought and heat stress, insect pests, early maturity, grain quality and dual-purpose grain-fodder performance (Boukar et al., 2016, 2020). Because these breeding priorities vary across environments, they strongly shape germplasm collection strategies; for example, breeding efforts focused on low-input resilience source germplasm from arid or marginal production regions. However, effective breeding requires a clear understanding of the breadth and structure of available genetic diversity, as this underpins the ability to identify novel alleles, incorporate underutilised variation, and assess the comparative value of material used across programmes (Boukar et al., 2019; Smith et al., 2015). Despite regional progress, the relationship among cowpea accessions from a comprehensive global geographic distribution remains poorly resolved. Although global core collections have been developed (Fatokun et al., 2018; Mahalakshmi et al., 2007; Muñoz-Amatriáin et al., 2021), most rely heavily on traditional cultivars from West Africa with limited representation from outside the continent, leaving gaps in understanding the full extent of global cowpea diversity. Integrating larger and more geographically diverse collections is therefore essential to capture the untapped genetic variation required to support breeding programmes worldwide and to position cowpea as a valuable genetic resource for global food and nutritional security.

Here, we explore a curated selection of the global gene pool of cowpea to understand patterns of genetic

diversity and population structure. The collection is particularly valuable because it integrates accessions from five genebanks and two collections, capturing a broad spectrum of geographic origins and genetic backgrounds. We assembled a dataset of 10 617 accessions, approximately 23% of the total global cowpea accessions, representing both core and diverse germplasm from six continents and 121 countries, and applied cost-effective and robust reduced-representation sequencing to generate high-quality genotypic data. The main objectives of this study were to (1) characterise a global collection of cowpea from seven international collections, (2) identify identical accessions within and between collections using genotypic data, and (3) assess the composition of the global cowpea collection in the context of diversity and genetic structure.

RESULTS

Metadata-based summary of the global cowpea collection (taxonomy and improvement status)

A total of 10 617 cultivated cowpea (*V. unguiculata*) accessions and one sister subspecies accession (*Vigna unguiculata* subsp. *stenophylla*) were used in this study and assessed with 4290 SNP markers (Figure S1). These accessions originated from diverse geographic regions and seven collections. The majority of these accessions were from the USDA Genebank (63.4%; $n = 6735$), followed by the IITA Genebank, Nigeria (18.3%; $n = 1942$), representing the core collection of cowpea accessions described by Mahalakshmi et al. (2007). The Australian Grains Genebank (AGG), Australia (7.4%; $n = 791$); the National Agriculture and Food Research Organization (NARO) Genebank, Japan (3.9%; $n = 414$); a collection from the University of California, Riverside (UCR), USA (3.3%; $n = 351$) representing the mini-core collection described by Muñoz-Amatriáin et al. (2021); a collection from the National Genebank of Mozambique at the Instituto de Investigação Agrária de Moçambique (IIAM; 2.6%; $n = 271$); and a Central and Southern American collection from Unidad de Genómica Avanzada, Langebio Cinvestav, Mexico (LAN; 1.1%; $n = 114$) also constituted the global collection.

The majority of cultivated cowpea assessed in this study belonged to the *unguiculata* cultivar group ($n = 5752$), with smaller representations from the *sesquipedalis* ($n = 240$) and *biflora* ($n = 36$) groups (Figure 1a). Forty-three per cent of the accessions had passport data at the *unguiculata* species ($n = 3670$) or subspecies ($n = 919$) level (Table S1). Cultivated cowpea accessions were classified into six improvement categories: Advanced/Improved Cultivar, Breeding Line, Landrace/Traditional Cultivar, Weedy, Wild and Unknown (Figure S2). The AGG collection consists primarily of Breeding Lines (74%), with a small percentage of Landrace/Traditional Cultivars (15%) and Unknown accessions (11%). In contrast, the IITA, IIAM,

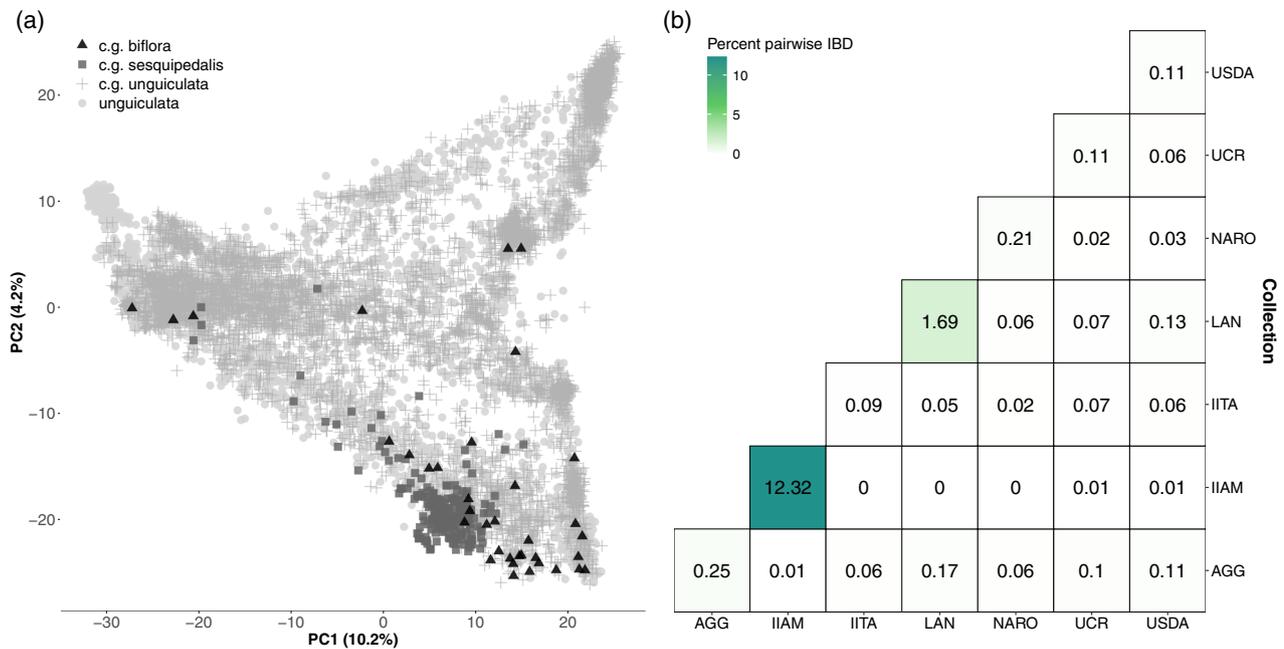


Figure 1. Genetic diversity of the cultivated global cowpea collection.

(a) Principal component (PC) analysis of 10 617 accessions colour-coded based on cultivar group (c.g.), where black triangles = *biflora* ($n=36$), grey squares = *sesquipedalis* ($n=240$), grey plus = *unguiculata* ($n=5752$) and grey circles = species level *Vigna unguiculata* or subspecies *unguiculata* combined ($n=4589$).

(b) Per cent pairwise identity by descent (IBD) greater than zero between collections. The percentage is calculated from the total number of pairs between each comparison compared with observed pairwise IBD values greater than zero.

NARO and UCR collections contain a large proportion of Landrace/Traditional Cultivars (86, 100, 79 and 74%, respectively). IITA and UCR also contain Breeding Lines (10 and 22%, respectively), while NARO also contains Unknown (12%) and Weedy (6%) accessions. The LAN collection contains mostly Unknown (98%) and Advanced/Improved Cultivars (2%). The USDA Genebank exhibited greater diversity, including Landraces (42%), Unknown (35%), Advanced/Improved Cultivars (19%) and Wild accessions (3%) (Table S2).

Identification of breeding material from genomic signatures

Pairwise identity by descent (IBD) was calculated for all 10 617 cultivated cowpea and assessed for differences between collections and geographic provenance. Pairwise IBD values ranged from 0 to 0.17% between collections and from 0.09 to 12.3% within collections (Figure 1b). Higher pairwise IBD was observed between AGG-LAN (0.17%), LAN-USDA (0.13%) and AGG-USDA (0.11%), while no pairwise IBD was detected between accessions IIAM-LAN, IIAM-IITA and IIAM-NARO. The highest pairwise IBD was found within the IIAM Genebank (12.3%), while all other collections had values below 2%. All geographic regions have an IBD peak around 0.95 and 1 (Figure 2a,b), while Southern Africa and the Americas have a larger IBD peak around 0.5 (Figure 2a,b).

The genomic relationship matrix heatmap (Figure 2c) provides a visual summary of related accessions with individuals colour-coded by geographic provenance in the columns and improvement status in the rows. As an example, a large group of accessions in the top left corner from Western Africa consisting of Advanced/Improved Cultivars, Breeding Lines, Landraces/Traditional Cultivars and Unknown accessions are similarly related. Another example is a small group of Landrace/Traditional Cultivars from Southern Africa in the bottom right corner that are highly related.

Duplication within and between collections

Potential duplicate accessions among 10 617 cultivated cowpea were identified using Nei's genetic distance with a threshold of ≤ 0.03 , Van Raden's genomic relatedness with a threshold of ≥ 1.3 and whole-genome IBD with a threshold of ≥ 0.6 . Each accession was compared with all other accessions in a pairwise manner for each of the metrics, resulting in a total of 112 710 072 pairwise combinations. Reciprocal pairs (e.g. Accession 1 versus Accession 2 and Accession 2 versus Accession 1) and identical pairs (e.g. Accession 1 versus Accession 1) were removed, leaving 56 355 036 unique combinations. Of these, 33 252 pairs met the three criteria and were classified as 'potentially duplicated accessions' (Table S3). A higher proportion of potential duplicates were detected within individual collections

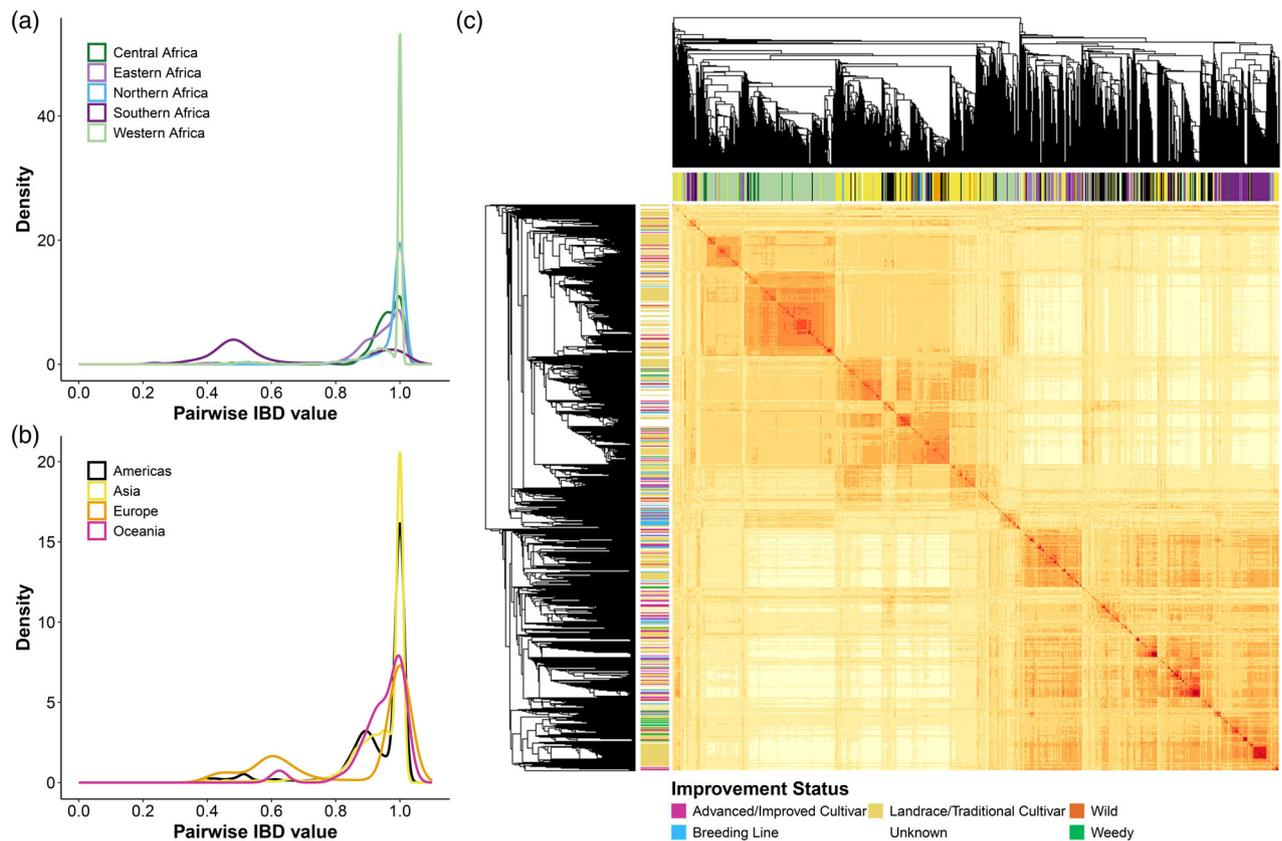


Figure 2. Genetic relatedness of the cultivated global cowpea collection.

(a, b) Pairwise identity by descent (IBD) frequency distribution for nine geographic regions.

(c) Genomic relationship matrix of 10 617 accessions, where yellow = dissimilar and red = similar. Accessions are coloured based on their geographic origin in the columns (colour key in [a] and [b]) and improvement status in the rows (colour key presented below the heatmap).

(57%) than between different collections (43%). Of these 33 252 pairs, 1329 were confirmed as ‘duplicated accessions’ based on passport alias names (Table S3). A higher proportion of duplicate accessions with matching names were detected between collections (74%) compared to within collections (26%). In cases where multiple pairs of accessions had an identical name, one or multiple accessions were retained depending on the genetic distance, GRM and IBD values. A total of 9609 non-redundant accessions were retained for subsequent analyses to assess population structure.

Global genetic structure of cowpea

Genetic structure

The genetic population structure of 9609 cultivated cowpea accessions in the non-redundant pruned dataset was inferred using three model-based clustering algorithms implemented in ADMIXTURE, STRUCTURE and LEA (Frichot & François, 2015). Cross-validation values from ADMIXTURE declined steadily with increasing K , without supporting a clear optimum, indicating the absence of a clearly defined optimal number of genetic groups in the

dataset. However, the log-likelihood and ΔK analyses indicated that $K=2$ showed the best-fit for both ADMIXTURE (Figure S3) and STRUCTURE (Figure S4). The sparse non-negative matrix factorisation (sNMF) cross-entropy analysis also supported two major populations, with further subdivision at higher K values (Figure S5). For all K values, an accession was assigned a genetic group if membership assignment to a group was greater than 70% (Table S4), while a lower percentage indicated admixture. At $K=2$, the structure analysis separated the cowpea germplasm into Population 1 and Population 2. The ADMIXTURE, STRUCTURE and sNMF results at $K=2$ were identical, so admixture proportions from sNMF are reported. The distribution of cowpea accessions between the two groups indicated that Population 2 (green) had the highest percentage of membership (43%) with 4115 accessions, while Population 1 (purple) had the lowest percentage of membership (35%) with 3415 accessions. However, the inferred ancestry indicated that 2079 accessions (22%) were admixed (Figure 3a,d, first panel). Support for $K=3$ indicated further division of the accessions into three populations. Population 1 remained as Population 1 while

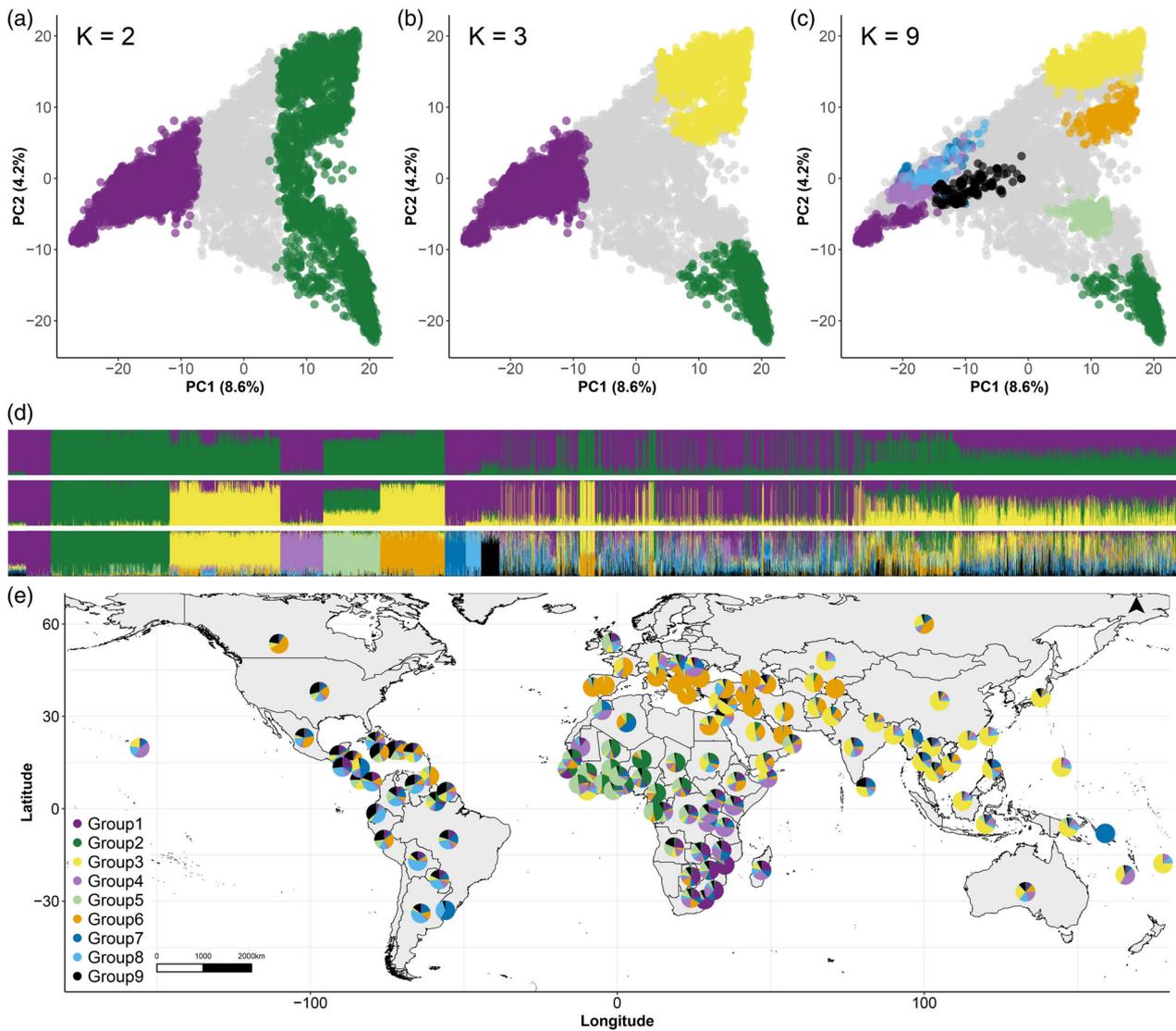


Figure 3. Population structure and geographic distribution of the non-redundant global cowpea collection. (a–c) Principal component (PC) analysis of the 9609 accessions coloured by population at $K=2$, $K=3$ and $K=9$. Colour key presented in (e). (d) Model-based clustering of $K=3$, $K=3$ and $K=9$. The x-axis has accessions ordered according to their assignment to each $K=9$ group. The y-axes quantify the membership to each ancestral population represented in colours corresponding to each $K=9$ group. Colour key presented in (e). (e) Geographic distribution of the genetic groups at $K=9$. For each country, the pie represents the proportion of the ancestral allele frequencies.

Population 2 was divided into two Populations: Population 2 and Population 3 (Figure 3b,d, second panel). Population 1 (purple) had the highest percentage of membership (34%) with 3275 accessions, Population 3 (yellow) had an intermediate percentage of membership (20%) with 1959 accessions, and Population 2 (green) had the lowest percentage of membership (13%) with 1266 accessions. Splitting the accessions into three populations resulted in an increase in the number of accessions admixed between populations ($n=3109$, 32%).

The populations at $K=2$ and $K=3$ reflect the broad genetic structuring in cowpea and have general geographic trends. At $K=2$, accessions belonging to Population 1

originate from Western Africa, Eastern Africa, Southern Africa and the Americas. Accessions in Population 2 originate from Western and Central Africa, Europe/Mediterranean and Asia (Figure S6). At $K=3$, when Population 2 was divided into Population 2 and Population 3, a clear distinction emerged, with African accessions in Population 2 and European/Mediterranean and Asian accessions in Population 3 (Figure S7).

We considered additional genetic structure to investigate the potential for population substructure across different regions. Support for a $K=9$ model further divided the accessions into nine genetic groups (Figure S5). Population 1 split into Groups 1, 4, 7, 8 and 9; Population 2 split into

Groups 2 and 5; and Population 3 split into Groups 3 and 6 (Figure 3c,d, third panel). Less than half of the 9609 accessions ($n = 4025$) were assigned to a group using a 70% membership threshold, with the remaining 5585 accessions admixed among groups (Figure S8).

Geographic enrichment of genetic groups

Geographic enrichment of the nine genetic groups was assessed using non-parametric tests, focusing on group proportions rather than individual assignments. Five groups exhibited significant geographic separation: Group 2 in Africa, Group 3 in Asia and Oceania, Group 6 in Europe and Group 7 in South America (Table 1). The remaining groups (Groups 1, 4, 5, 8 and 9) did not reach statistical significance but showed the highest representation in specific regions: Group 1 in Southern Africa, Group 4 in Eastern Africa, Group 5 in Central and West Africa, Group 8 in South and Central America, and Group 9 in the Americas (North America, Central America, Southern America and the West Indies) (Table S5). To address regional sample size bias, each geographic region was rarefied to the same sample size ($n = 50$) and tested for significance over 500 permutations. Results from the rarefaction analysis were consistent with those obtained from the full dataset (Table S6). These results highlight both broad-scale and region-specific geographic patterns among cultivated cowpea, reflecting historical dispersal patterns.

Phylogeny of cowpea

The cowpea phylogenetic tree (Figure 4) is well supported, with all 9607 nodes showing bootstrap (BS) values $>50\%$ and a mean of 80.7% (Table S7). Support is highest for small terminal groups (BS 90–100%), whereas backbone edges are shorter and moderately supported (mean BS 64%). A total of 34 clades were identified along the backbone of the tree (Table S7), 12 of which contain over 200 accessions, including seven clades with more than 500 accessions. A subset of these large clades contains accessions that generally align with the populations, groups and have trends for improvement status, geographic provenance and cultivar group (Figure S9). A summary of the clades within this tree and detailed historic, taxonomic and geographic patterns can be found in Appendix S1, Figures S10–S17 and Table S7.

Genetic diversity and relationships between genetic groups

Analysis of molecular variance (AMOVA) was computed between the populations and groups excluding the mixed accessions. Total molecular variance was partitioned into 32% between populations and 68% within populations at $K = 2$. At $K = 3$, total molecular variance was partitioned into 37% between populations and 63% within populations. At

Table 1 Non-parametric assessment of geographic enrichment of group proportions for the nine genetic groups

Region	<i>n</i>	Group 1			Group 2			Group 3			Group 4			Group 5			Group 6			Group 7			Group 8			Group 9		
		Mean	Sig		Mean	Sig		Mean	Sig		Mean	Sig		Mean	Sig		Mean	Sig		Mean	Sig		Mean	Sig		Mean	Sig	
Africa	5125	0.152	a		0.272	a		0.041	b		0.114	b		0.154	a		0.088	e		0.075	bc		0.045	e		0.059	c	
Asia	2654	0.042	d		0.017	d		0.429	a		0.117	b		0.026	c		0.134	c		0.115	b		0.073	d		0.046	d	
Central America	82	0.077	abc		0.009	bcd		0.103	b		0.084	bd		0.067	abc		0.122	bcde		0.076	bc		0.232	ab		0.230	ab	
Europe	255	0.019	f		0.017	bc		0.107	bc		0.032	e		0.032	d		0.678	a		0.065	d		0.033	f		0.016	e	
North America	968	0.039	e		0.017	c		0.060	c		0.066	c		0.064	b		0.222	b		0.096	c		0.151	c		0.284	a	
Oceania	68	0.050	acd		0.018	bd		0.322	a		0.184	a		0.038	bcd		0.098	de		0.055	bc		0.143	bc		0.092	cd	
South America	277	0.100	b		0.020	bd		0.044	b		0.078	ad		0.052	b		0.068	e		0.156	a		0.280	a		0.202	a	
Unknown	180	0.053	cde		0.025	bc		0.133	b		0.096	bc		0.046	bc		0.195	bcd		0.102	bc		0.178	bc		0.173	b	

Note: Mean group proportions for each region are presented with compact letter display indicating a significant difference in means across regions. Significant group enrichment for geographic regions are italicised. North and Central America are separated into two regions. *n* = Geographic region sample size.

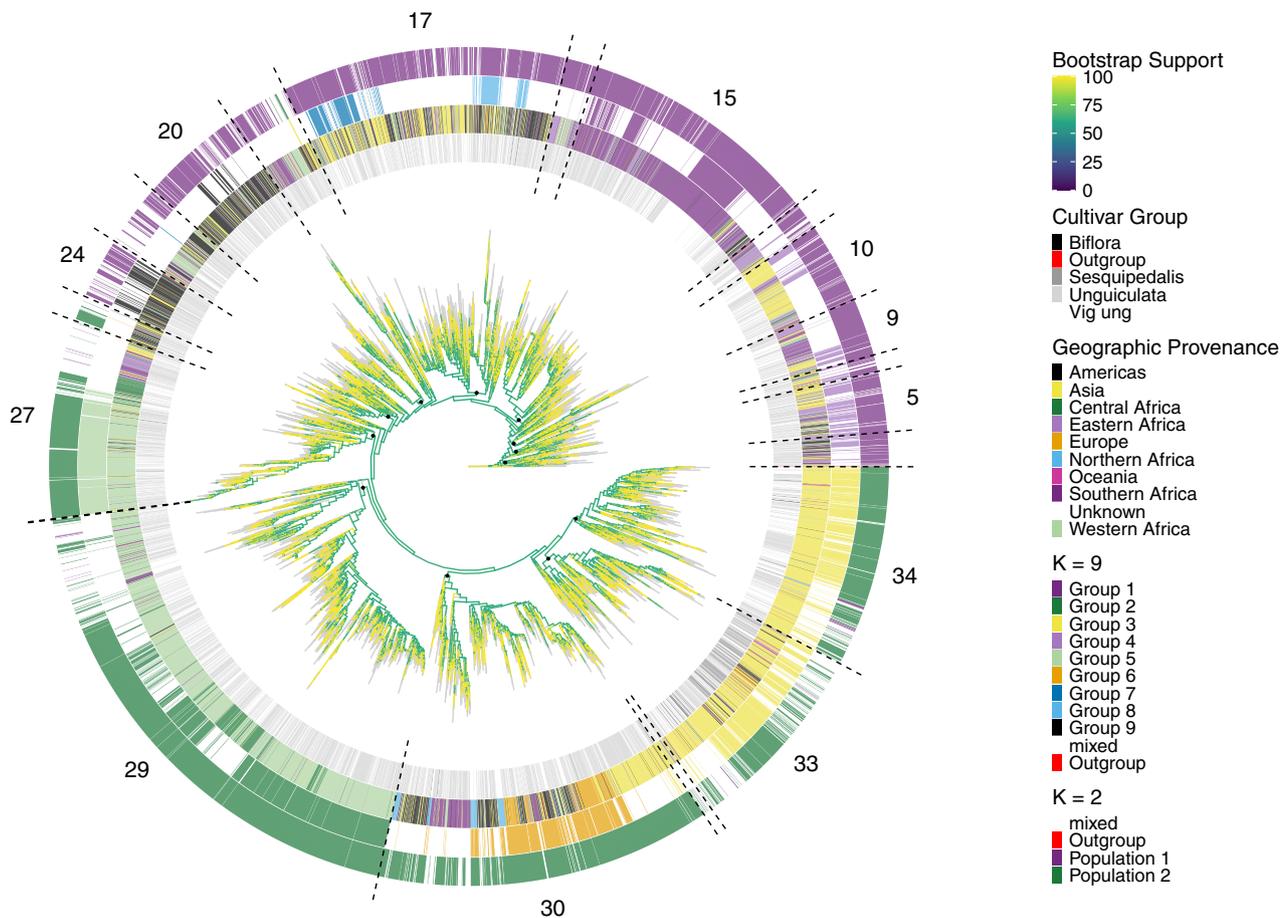


Figure 4. Maximum likelihood phylogeny of 9610 cowpea accessions implemented in IQ-TREE.

Bootstrap support values for each node are shown as a continuous colour scale on the branches from 0% in dark blue to 100% in yellow, while grey indicates the branches leading to the tips of the tree. Coloured bars on the outside of the tree ordered from the centre ring outwards show: the cultivar group, geographic provenance, genetic group at $K=9$ and population at $K=2$ for each accession. The 12 largest clades are numbered outside the tree with dashed lines indicating clade boundaries and black circles indicating their common ancestral node. See Table S7 for accession names and their corresponding clade.

$K=9$, total molecular variance was partitioned into 50% between groups and 50% within groups (Table 2). Although lower percentages of variation were found between groups at all hierarchy levels, all variance partitioning was statistically significant ($P < 0.01$). The analyses indicated high variance prevailing within the groups identified. AMOVA was also computed between collections, and most molecular variation was partitioned within collections (95%) compared to between collections (5%; Table 2).

Diversity metrics including the number of private alleles, allelic richness, nucleotide diversity, inbreeding coefficient (F_{IS}), IBD, observed heterozygosity and the number of private alleles were assessed for each of the nine groups. Groups 1 (Southern Africa), 2 (Western/Central Africa) and 3 (Asia) contained the most private alleles ($n=1998-4641$; Table S8) compared with Groups 4 (Eastern Africa), 5 (Central/Western Africa), 6 (Europe), 7 (South America), 8 (South/Central America) and 9 (Americas) ($n=0-656$). Group 4 contained the most

allelic richness (0.182) and nucleotide diversity (0.16) but a relatively high inbreeding coefficient (0.93) and few private alleles ($n=656$; Table 3). Groups 1, 3 and 4 had large π values compared with Groups 7, 6 and 8. Groups 3–5 were the most inbred compared with Groups 1, 6–8 and Group 5 had the lowest observed heterozygosity ($H_O=0.007$) (Table 3). Group 1 had the lowest inbreeding coefficient ($F_{IS}=0.77$; Table 3), the largest observed heterozygosity ($H_O=0.033$) and a large proportion of breeding material (i.e. $\hat{\pi}$ of approximately 0.5 is indicative of parent/child relationships and sib-pairs; Figure 5a). Africa was the only continent to contain private alleles (Table S9) and had the largest nucleotide diversity and allelic richness (Table S10).

Pairwise F_{ST} , or the population stratification statistic, was estimated between the $K=9$ groups (Figure 5b). The highest genetic differentiation of 0.73 was found between Group 6 (Mediterranean) and Group 7 (South American). Similar high differentiation of 0.7 was also observed between Group 6 and Group 8 (Central and South

Table 2 Analysis of molecular variance (AMOVA) partitioning the variance between and within genetic groups and collection

Hierarchy level	Source of variation	d.f.	SS	MS	P-value	Variance (%)
Population ($K=2$)	Between populations	1	975 285.8	975 285.8019	<0.001	31.95
	Within populations	7528	4 186 342.2	556.1028		68.05
	Total	7529	5 161 628	685.5662		100
Population ($K=3$)	Between populations	2	1 198 696	599 347.781	<0.001	36.74
	Within populations	6497	3 338 295	513.821		63.26
	Total	6499	4 536 991	698.106		100
Group ($K=9$)	Between groups	8	1 205 494	150 686.7596	<0.001	50.08
	Within group	4015	1 423 063	354.4365		49.92
	Total	4023	2 628 557	653.3822		100
Collection	Between collections	6	200 120.5	33 353.4246	<0.001	5.32
	Within collection	9602	6 346 552.6	660.9615		94.68
	Total	9608	6 546 673.2	681.3773		100

Note: P-value calculated from 999 permutations.

Abbreviations: d.f., degrees of freedom; MS, mean square; SS, sum of squares.

Table 3 Diversity metrics of the nine genetic groups

Group	Allelic richness	Private alleles	π	F_{IS}	H_E	H_O
Group 1 (Purple)	1.72	2089	0.147	0.774	0.147	0.033
Group 2 (Green)	1.69	4641	0.104	0.907	0.104	0.010
Group 3 (Yellow)	1.74	1998	0.158	0.934	0.158	0.011
Group 4 (Light Purple)	1.82	656	0.161	0.932	0.160	0.011
Group 5 (Light Green)	1.60	630	0.108	0.937	0.108	0.007
Group 6 (Orange)	1.68	8	0.078	0.872	0.078	0.010
Group 7 (Blue)	1.73	0	0.065	0.846	0.065	0.010
Group 8 (Light Blue)	1.69	9	0.081	0.846	0.081	0.012
Group 9 (Black)	1.63	0	0.134	0.924	0.134	0.010

Abbreviations: π , weighted mean nucleotide diversity; F_{IS} , weighted mean inbreeding coefficient; H_E , expected heterozygosity; H_O , observed heterozygosity.

America). High pairwise F_{ST} values of 0.679 and 0.658 also indicated that Group 7 and Group 8 were distinct from Group 2 (West African). In contrast, the lowest genetic differentiation of 0.289 was found between Group 6 and Group 3 (Asian).

DISCUSSION

The purpose of this study was to examine population structure and diversity of cultivated cowpea sourced from multiple genebanks and collections. The accessions in this collection reflect the global genetic diversity of extant cowpea. This dataset includes the largest cohort of cowpea genotyped using a single technology to date. The collection includes previously uncharacterised accessions and integrates previously characterised collections, creating a valuable resource for cowpea research and crop improvement.

Geographic structure of genetic diversity mirrors historical dispersal

Cowpea was domesticated in sub-Saharan Africa and subsequently spread across the Mediterranean Basin, Asian

and the Americas (Herniter et al., 2020; Muñoz-Amatriáin et al., 2021), creating multiple opportunities for founder effects, drift and selection under new agronomic and environmental conditions. As a result, the contemporary genetic structure of cowpea is expected to retain a geographic 'footprint' of historical dispersal routes and region-specific improvement, an expectation supported by prior population genetic and genomic surveys of global and curated diversity sets (Fiscus et al., 2024; Huynh et al., 2013; Muñoz-Amatriáin et al., 2021; Xiong et al., 2016). In particular, archaeological (D'andrea et al., 2007; Goldstein et al., 2024), textual (Herniter et al., 2020) and genetic evidence (Huynh et al., 2013; Xiong et al., 2016) suggests at least two major dispersal routes: one from Africa into the Old World (Mediterranean, India and Southeast Asia) (Fuller, 2003) and another into the New World during the Columbian Exchange (Herniter et al., 2020), consistent with the recurring observation that non-African cowpea often clusters with distinct African regions (Fiscus et al., 2024; Herniter et al., 2020; Huynh et al., 2013; Muñoz-Amatriáin et al., 2021; Xiong et al., 2016).

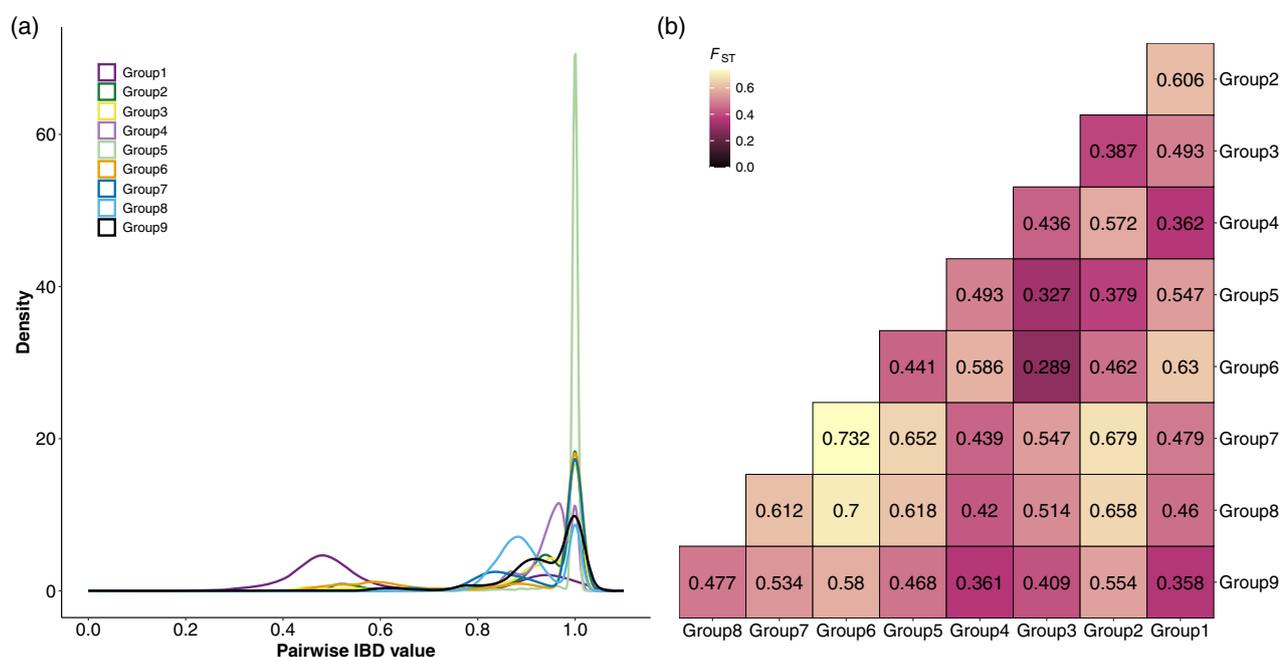


Figure 5. Population relatedness of the non-redundant global cowpea collection.

(a) Pairwise identity by descent (IBD) frequency distribution for each genetic group at $K=9$.

(b) Pairwise F_{ST} between genetic groups at $K=9$, coloured continuously from low differentiation (black) to high differentiation (yellow).

A substantial body of work supports West Africa as a major centre of diversity for cultivated cowpea and identifies broad geographic partitioning among African gene pools, separating Western and Central African material and Eastern and Southern African material, and for non-African accessions showing affinities to one or the other (Fatokun et al., 2018; Huynh et al., 2013; Padulosi & Ng, 1997; Steele, 1976). Huynh et al. (2013) identified two major African gene pools and showed that Asian and European accessions tend to share closer affinity with West African material, whereas many American accessions align more closely with East African material, suggesting a pattern consistent with historically distinct introduction routes and subsequent local diversification. Our results broadly agree with this framework: the two primary clusters detected in model-based analyses are consistent with a deep, geographically structured partitioning of cowpea diversity, while finer structure reveals additional subdivisions that plausibly reflect multiple introductions, subsequent drift and selection during regional cultivation and breeding. This interpretation is also consistent with analyses of curated resources (Muñoz-Amatriáin et al., 2021), which recover geographically patterned subpopulations and highlight the contribution of West African diversity to multiple lineages worldwide.

Interpreting the number of genetic clusters (K) requires explicit caution because STRUCTURE-like models infer ancestry under assumptions (e.g. Hardy–Weinberg equilibrium within clusters, linkage equilibrium among

loci) and require *post hoc* evaluation of K , which can be sensitive to both hierarchy and sampling design (Pritchard et al., 2000). The widely used delta K approach is known to preferentially identify the uppermost hierarchical level of structure and can show a bias towards $K=2$ in many realistic scenarios, meaning that a ‘two-population’ solution can represent an informative top-level partition without capturing finer historical complexity (Janes et al., 2017) and should be interpreted as a broad summary rather than the full complexity of cowpea’s global population history. Uneven sampling across regions and collections can further distort inference of K and ancestry proportions and has been highlighted as a limitation for STRUCTURE-based interpretation unless accompanied by subsampling or sensitivity analyses. Accordingly, the additional structure inferred here using sNMF avoids assumptions of no genetic drift, Hardy–Weinberg equilibrium and linkage equilibrium in ancestral populations and is more appropriate to deal with inbred lineages (Frichot et al., 2014). Nevertheless, given the strong imbalance in sampling intensity among regions and collections, inferred groups and private-allele patterns should be treated as indicative rather than definitive although rarefaction-based geographic analyses were supportive of geographic enrichment of genetic groups (Table S6).

Phylogenetic reconstruction provides complimentary insight into relationships among accessions, but the short backbone and moderate support at deeper nodes observed here likely reflect a rapid radiation from Africa in which

multiple lineages diverged over short time intervals, leaving limited phylogenetic signal for resolving basal relationships (Whitfield & Lockhart, 2007). In such contexts, incomplete lineage sorting and historical gene flow can generate substantial gene-tree discordance (Degnan & Rosenberg, 2009; Hibbins & Hahn, 2021), and strictly bifurcating trees may inadequately represent evolutionary history (Bjornson et al., 2024), particularly in domesticated crops where introgression, and breeding create reticulate patterns (Degnan & Rosenberg, 2009; Hibbins & Hahn, 2021; Mallet, 2005). These processes complicate the reconstruction of deep evolutionary histories in rapid radiations, motivating the use of genome-wide genetic datasets, coalescent-aware inference and network-based representations that better accommodate admixture (Degnan & Rosenberg, 2009; Hibbins & Hahn, 2021; Liu et al., 2014; Philippe et al., 2011; Whitfield & Lockhart, 2007). Despite limited resolution at deep nodes, the cowpea phylogeny remains informative for more recent divergences, particularly among well-supported smaller clades and clusters of closely related accessions, which are directly relevant for germplasm management (e.g. identifying clusters of closely related accessions) and for guiding the choice of diverse parents in breeding.

The strong relationship between inferred genetic groups and geographic origin indicates that spatial factors have significantly influenced the genetic differentiation among accessions and reinforces the value of geographically diverse sampling for both evolutionary inference and applied utilisation. From a conservation and breeding perspective, geographically structured diversity highlights reservoirs of alleles likely shaped by regional selection pressures and improvement histories, providing opportunities to prioritise underrepresented regions and guide the selection of accessions for breeding programmes aimed at improving crop performance under region-specific constraints. To further resolve domestication history and inter-continental dispersal, future work integrating representative sampling of wild cowpea, plastome phylogenies and haplotype analysis of domestication-related loci would provide stronger tests of single versus multiple domestication hypotheses and clarify the contribution of post-domestication gene flow.

Modern germplasm sharing and implications for cowpea conservation and management

Global sharing of cowpea germplasm has increased access to diverse genetic resources, but it also promotes redundancy through repeated collection and redistribution of identical or near-identical material. In this global dataset, we identified 1008 duplicate accessions (9.5%), indicating that duplication is present but comparatively lower than reported in other crops, such as: barley, cassava and soybean which show 33% (Milner et al., 2019), 26%

(Albuquerque et al., 2019) and 23% (Song et al., 2015) duplication, respectively. Differences in duplication rates among crops likely reflect a combination of collection history, naming or curation practices, and the resolution of genetic data used for detection, in addition to crop biology. Importantly, our estimate is likely conservative because the present study does not encompass all globally available accessions. Genesys currently holds approximately 45 500 amalgamated cowpea accessions (Data accessed through Genesys, <https://www.genesys-pgr.org/c/cowpea>, 2 October 2025), and several major collections, such as the IITA (Ibadan, Nigeria, approximately 19 000 accessions), the NBPGR (New Delhi, India, approximately 3700 accessions) and the full collection held at the UCR (CA, USA, approximately 6000 accessions), contain substantial material not fully represented here. If these additional approximately 28 000 accessions were included in analyses, we would expect duplication within and among collections to increase, reinforcing the need for coordinated curation approaches that distinguish truly unique accessions from redundant entries.

Patterns of duplication provide insight into the processes generating redundancy. Accessions in this study originated from more than 120 countries, and re-collecting from the same locations, combined with routine germplasm exchange between institutions, can readily produce duplicated entries across genebanks. Most duplicates were detected between collections (74%), indicating material exchange accompanied with appropriate documentation. However, the large number of molecularly similar pairs (33 252) highlights an additional challenge: high genetic similarity can reflect true duplication, but it can also reflect breeding relatedness (e.g. shared parentage, backcrossing to introduce simply inherited traits, or repeated use of elite lines in breeding programmes as noted by Boukar et al. (2020)). In practice, both processes likely contribute. This distinction matters operationally, because redundancy inflates curation costs and can lead to users unknowingly select duplicates if accessions are sampled randomly. While closely related breeding materials may remain valuable, appropriate pedigree, passport and standardised trait metadata will greatly enhance and support downstream use.

Limitations in passport metadata further complicate both management and biological interpretation. Only 4113 accessions in this global collection were geo-referenced or resolved beyond country level (Table S1), constraining fine-scale geographic inference and reducing our ability to link genetic patterns to specific environments. However, this limitation can be partially mitigated by integrating genetic data with a Focused Identification of Germplasm Strategy (FIGS) (Sunitha et al., 2024). By linking poorly annotated accessions to genetically similar accessions (Tables S3 and S11) with high-quality collection site

information, FIGS provides a practical path to infer likely adaptive traits and prioritise subsets of germplasm for evaluation (Khazaei et al., 2013). This integrative approach would enable breeders and users to move beyond broad or random screening towards more targeted identification of candidate accessions likely to contain adaptive or agronomic traits aligned with specific breeding objectives.

From a management perspective, this study suggests that substantial molecular variance is retained within collections, rather than between (Table 2), indicating that local collections contain distinct germplasm, supporting continued investment in regional collections and emphasising the value of coordinated global curation. Given declining sequencing and genotyping costs, a feasible long-term goal is a globally coordinated genotyping-based curation effort across genebanks to identify unique accessions, flag candidate duplicates and standardise identity tracking across genebanks (Anglin et al., 2025). In this context, this study demonstrates that low-density SNP markers can serve as an efficient initial screening tool to identify candidate duplicate sets for curator review. These candidates can then be prioritised for confirmatory evaluation using higher density genotyping and/or phenotypic evaluation (Song et al., 2015). Importantly, new field collections should be genotyped and assessed for similarity to pre-existing accessions to prevent accumulated redundancy and support strategic decisions about which accessions should be prioritised for additional back up of irreplaceable accessions.

IBD analyses further supports the interpretation that both true duplication and pedigree-related relationships contribute to the observed patterns. The strong IBD peak around 1 is consistent with many similar and near-identical accessions (Figure 2a,b), while the IBD peak around 0.5 in regions such as Southern Africa and the Americas (Figure 2a,b) suggests large contributions of breeding material (parent/offspring and half-sib relationships) originating from these regions. Several limitations should be acknowledged. First, within-accession heterogeneity was not assessed; although cowpea is predominantly selfing with low outcrossing rates of 0–5% (Fatokun & Ng, 2007; Gumede et al., 2022; Xiong et al., 2016), within-accession heterogeneity could still influence pairwise similarity and duplicate detection in some cases. Second, duplication inference is constrained by marker density: while the 3073 SNPs used here are sufficient for broad screening, structure and diversity estimates, they may fail to resolve very closely related material from non-identical material. Therefore, the candidate duplicate list should be treated as a prioritised set for germplasm management follow-up rather than a definitive catalogue. Corroborating evidence for shared origin can be sought through (1) comparison of vouchers in the genebanks' herbarium, (2) inspection of the original (paper) passport records, (3) assessing higher density marker data for a larger number of individuals

from candidate duplicates and (4) field evaluation in which candidate duplicates are grown side by side. Although these steps are labour intensive, the upfront investment is likely to be outweighed by the long-term cost of maintaining superfluous material and the downstream inefficiencies created by redundant sampling.

Taken together, these findings emphasise that effective cowpea germplasm conservation now requires both genetic curation and standardised information management (Milner et al., 2019). Improving the consistency of passport descriptors and standardising trait and pedigree data, supported by ongoing digital efforts across genebanks, will significantly enhance the accuracy, accessibility and long-term value of germplasm collections. In parallel, genotype-informed curation provides a framework for reducing redundancy, prioritising globally unique accessions for long-term conservation and safety duplication, and enabling breeders to more efficiently identify germplasm with likely adaptive potential.

MATERIALS AND METHODS

Plant materials

A total of 10 617 cultivated cowpea (*V. unguiculata*) accessions and one sister subspecies accession (*V. unguiculata* subsp. *stenophylla*) were used in this study. These accessions originated from diverse geographic regions and seven collections. The majority of these accessions were from the USDA Genebank (63%; $n = 6735$), followed by the IITA Genebank, Nigeria (18%; $n = 1942$), representing the core collection of cowpea accessions described by Mahalakshmi et al. (2007). A collection of accessions held at AGG, Australia (7%; $n = 791$) and the NARO Genebank, Japan (4%; $n = 414$), were included to expand geographic distribution. The diversity of the cowpea accessions was further enriched by the inclusion of several smaller but highly valuable collections from various regions. These included 351 accessions from the UCR mini-core collection (Muñoz-Amatrián et al., 2021), 271 accessions from the National Genebank of Mozambique at the IIAM (Macharia et al., 2025) and 114 accessions from the Central and Southern American collection from Unidad de Genómica Avanzada, LAN. Together, these diverse sources represent a broad spectrum of genetic variation, including both regional and global cowpea germplasm, thereby providing a comprehensive foundation for subsequent genetic studies. Passport descriptors (i.e. metadata) including taxonomic classification, location of provenance, collection site and improvement status were collated for all accessions from their respective collection, where available. Metadata were accessed online through: the USDA-ARS Germplasm Resources Information Network (GRIN) (<https://www.ars-grin.gov/>), IITA Genebank (<https://my.ita.org/accession2/browse.jsp>), AGG GRIN-Global Database System (<https://ausgenebank.agriculture.vic.gov.au/gringlobal/search>), NARO (https://www.geneaffrc.go.jp/databases-plant_search_en.php), Genesys (<https://www.genesys-pgr.org/>) and Supporting Information from Muñoz-Amatrián et al. (2021) and Macharia et al. (2025). Metadata for accessions used are provided in Table S1.

Genotyping-by-sequencing

Genotyping of the global cowpea collection was conducted by Diversity Arrays Technology (DArT, Canberra, ACT, Australia)

Pty Ltd (<https://www.diversityarrays.com/services/dartseq/>) using the DArTseq™ reduced-representation sequencing platform (Kilian et al., 2012). For each accession, 30–40 mg of seed tissue derived from single or multiple seeds was sampled and placed in a 96-well plate. Samples were ground in a TissueLyser II (QIAGEN, Hilden, Germany) using metal rods at 30 Hz for 20–30 sec, repeated twice. Following grinding, the rods were removed and tubes sealed. Ground samples were stored at 4°C before shipping to DArT where DNA extraction was conducted using a modified cetyl trimethyl ammonium bromide (CTAB) method (Doyle & Doyle, 1987). For accessions restricted by quarantine regulations, leaf material from the NARO accessions were sampled following the procedure described by Ofem et al. (2025) and sent to DArT for DNA extraction. Samples from the IIAM collection were extracted at the Genbank in Mozambique using the protocol described by Macharia et al. (2025) and subsequently shipped to DArT for sequencing. All DNA samples were then digested with methylation sensitive restriction enzymes (*Pst*I and *Mse*I) to reduce genome complexity and to remove repetitive sequences (Kilian et al., 2012). Libraries with size lengths of 200–500 bp were constructed using a Truseq-Nano DNA HT Sample Preparation Kit (Illumina; catalogue no. FC-121-4003) following the manufacturer's recommendations and sequenced on an Illumina HiSeq 2500 platform to produce 77 bp paired end reads (Sansaloni et al., 2011). Raw sequencing reads were processed using DArT's proprietary analytical pipeline. In this pipeline, the FASTQ files were filtered to remove adapters, low-quality reads and redundant reads with barcode regions stringently filtered (Phred pass score ≥ 30) to ensure accurate sample assignment to the correct reads. Identical reads were collapsed into 'fastqcall' tag files which were then used as input for DArT's secondary pipeline for SNP analysis (Raman et al., 2014). Reads were subsequently aligned to the cowpea reference genome IT97K-499-35 v1.2 (Lonardi et al., 2019) using BWA-MEM v0.7.17 (Li & Durbin, 2009), and SNP genotypes were called with DArT's proprietary secondary pipeline (DArTsoft14) (Sansaloni et al., 2020). Genotype data were delivered in VCF format (Cruz et al., 2013). During post-DArT quality control, samples with a call rate ≤ 0.50 and markers with a call rate ≤ 0.80 were excluded. The overall call rate in the raw genotypic dataset was approximately 82%. Missing genotypes were inferred using Beagle v 5.0 (Browning & Browning, 2016). Further filtering was applied by removing markers with heterozygosity ≥ 0.4 and individuals with heterozygosity ≥ 0.2 . The final filtered dataset consisted of 10 618 accessions and 4290 high-quality SNPs, denoted the 'raw dataset', providing a robust dataset for subsequent analyses of genetic diversity. The genomic distribution and density of these SNPs were visualised using the online SNP Density tool (<https://www.bioinformatics.com.cn/en>) (Figure S1).

Genetic relatedness between collections and dataset generation

Cultivated cowpea were extracted from the raw dataset and markers with a minor allele frequency ≥ 0.01 were retained using VCFtools v 0.1.16 (Danecek et al., 2011), resulting in 3073 SNPs across 10 617 accessions denoted the 'cultivated dataset'. An initial principal component analysis (PCA) was performed using the *prcomp* function from the stats package in R v 4.3.0 (R Core Team, 2023) to visualise accession distribution according to their metadata, including cultivar group, improvement status, collection (i.e. germplasm source) and geographic provenance.

Duplicate accessions within the cultivated dataset were identified based on three metrics: genetic distance, genomic relatedness and IBD. Pairwise Nei's genetic distance was calculated

using the *nei.dist* function from the poppr package v 2.9.4 (Kamvar et al., 2014), ranging from 0 (minimal genetic difference) to 1 (maximal genetic difference) using the cultivated dataset. A conservative threshold of ≤ 0.03 was used to identify genetically similar pairs. Van Raden's genomic relationship matrix was calculated using the *Gmatrix* function from the AGHmatrix package v 2.1.4 (Amadeu et al., 2023) using the cultivated dataset, with pairwise values ≥ 1.3 considered indicative of genetic relatedness. Linkage disequilibrium (LD) pruning was performed using PLINK v 1.90b7.2 (Purcell et al., 2007) on the cultivated dataset with the parameter *--indep-pairwise 50-5-0.5* to generate the 'cultivated-pruned dataset', consisting of 2302 unlinked SNPs. This dataset was used to calculate identify by state (IBS), defined as the average proportion of alleles shared between pairs of individuals at genotyped SNPs. The proportion of whole-genome IBD was also calculated using the *--genome* flag and PLINK's Hidden Markov Models (HMM) algorithm (Purcell et al., 2007). Accessions with a $\hat{\pi}$ value greater than 0.5 indicated parent/child relationships, sib-pairs and duplicated accessions. Accession pairs would need to fit all three criteria to be considered a potential duplicate: genetic distance ≥ 0.03 , genomic relatedness ≥ 1.3 and $\hat{\pi} \geq 0.6$. To adopt a more conservative approach, passport metadata were also considered: accession pairs that met all three genetic criteria were required to share the same accession name or alias. Potentially duplicate accessions with shared names were thus identified as duplicates and removed from the raw dataset using bcftools v 1.15.1 (Danecek et al., 2021). This dataset was filtered to remove markers with a minor allele frequency ≤ 0.01 resulting in 3083 SNPs and 9609 accessions and was denoted the 'non-redundant dataset'. The non-redundant dataset was also LD pruned using the methods described above, resulting in 2312 SNP retained to generate the 'non-redundant pruned dataset'.

In the cultivated-pruned dataset, relatedness was summarised as the proportion of accession pairs with $\hat{\pi} > 0$ within and between collections, reported as 'per cent pairwise IBD' in Figure 1b. Relatedness was calculated for all pairs between 10 617 accessions using PLINK v 1.90b7.2 (Purcell et al., 2007) by summing the pairs with $\hat{\pi} > 0$ within a collection combination, then calculating the total number of pairs within the collection combination and finally calculating the percentage of related pairs relative to the total possible combinations for the collection combination. In addition, the distribution of $\hat{\pi}$ values within each geographic provenance was assessed with density plots in R, reported as 'pairwise IBD values' in Figure 2a,b.

Genetic structure and diversity analysis

To explore the overall patterns of genetic clustering in cowpea germplasm, two separate analyses were performed: (1) Model-based structure analyses used ADMIXTURE v 1.3.0 (Alexander et al., 2009), STRUCTURE v 2.3.4 (Pritchard et al., 2000) and the sNMF algorithm (Frichot et al., 2014) implemented in the LEA package v 3.12.2 (Frichot & François, 2015); and (2) Phylogeny reconstruction used IQ-TREE v 2.2.2.3 (Minh et al., 2020). The 'non-redundant pruned dataset' described above was used for structure analyses and phylogenetic analyses.

ADMIXTURE, STRUCTURE and sNMF were used to perform model-based clustering analyses to explore the broad patterns of population structure and determine the number of putative ancestral populations (*K*). STRUCTURE uses a Bayesian clustering approach, ADMIXTURE uses maximum likelihood modelling, and sNMF uses sparse nonnegative matrix factorization algorithms to estimate individual ancestry coefficients. As ADMIXTURE is computationally faster to run compared with STRUCTURE, the

ADMIXTURE analysis was performed using 10 independent runs specified with a different seed using `-s` with each run set from 1 to 40 with 1- cross-validations (`-cv = 10`) to determine the K values with the lowest cross-validation error. As the cross-validation error continued to decline as K increased, the optimal K value was determined using the delta K method (Evanno et al., 2005). Subsequently, 10 independent runs for each simulated value of K ranging from 1 to 15 was performed using STRUCTURE. Each run consisted of a burn-in period of 50 000 iterations, followed by 100 000 Markov Chain Monte Carlo (MCMC) replications. The putative optimal K was determined using the `pophelperShiny` app v 2.1.1 (Francis, 2017) using the delta K method (Evanno et al., 2005). In parallel, model-based estimation of ancestral coefficients was performed using the sNMF algorithm (Frichot et al., 2014) implemented in the R package LEA (Frichot & François, 2015). The number of ancestral populations (K) was set from 1 to 25 with 20 iterations for each K value. The best-fit K value was determined by assessing the change in cross-entropy for each K , that is the elbow of the curve. Ancestral proportions from the sNMF analysis were plotted using the `PophelperShiny` app v 2.1.1 (Francis, 2017). Accessions were assigned to populations if their individual admixture coefficient matrix (Q) was at least 70%, and those with less than 70% membership probability were considered mixed. Population structure was further examined using the `prcomp` function to perform a PCA. Distribution of genetic groups to country of provenance was visualised in R using the `worldmap` v 1.3-8 (South, 2011) and `ggplot2` v 3.5.1 packages (Wickham, 2016).

For the phylogeny reconstruction, one *V. unguiculata* subsp. *stenophylla* accession was included as the outgroup in the non-redundant pruned dataset. A maximum likelihood phylogeny was generated using IQ-TREE v 2.2.2.3 (Minh et al., 2020) specifying the polymorphisms-aware phylogenetic model (Schrempf et al., 2016, 2019). ModelFinder (Kalyaanamoorthy et al., 2017) was first used to determine the best-fit substitution model for the data (GTR+F+ASC+R10). Default parameters were used for tree construction with the *stenophylla* accession assigned as the outgroup. Convergence was not assessed by undertaking independent runs of IQ-TREE and examining tree topology due to the large number of tips (9610) and nodes, but did ensure that the log-likelihood values were stable at the end of the run. Branch support was performed using 10 000 replicates of UFboot (Hoang et al., 2018). The 9610 taxa tree was plotted in R using `ggtree` v 3.8.2 (Yu et al., 2017) and `ggtreeExtra` v 1.10.0 (Xu et al., 2021).

Relatedness in terms of the proportion of IBD was calculated for all pairwise accessions using the 'non-redundant pruned dataset' and the distribution of these proportions were assessed within each group using a density plot in R. To assess genetic diversity, the non-redundant dataset was used to estimate six diversity metrics. Nucleotide diversity (π ; Hohenlohe et al., 2010), F_{IS} (Weir & Cockerham, 1984), pairwise F_{ST} (Weir & Cockerham, 1984), expected and observed heterozygosity (H_E and H_O , respectively) were calculated for each of the genetic groups and eight geographic regions using `snpr` v 1.2.9.2 (Hemstrom & Jones, 2023). Allelic richness (Hurlbert, 1971) and the number of private alleles for groups and regions were calculated using `hierfstat` v 0.5-11 (Goudet, 2005) and `poppr` v 2.9.4 (Kamvar et al., 2014, 2015), respectively. AMOVA was used to partition the total variance in allele frequencies within and among hierarchies to assess the significance level at each partition: $K=2$, $K=3$, $K=9$ and collection, and was implemented in the packages `poppr` v2.9.4 (Kamvar et al., 2014, 2015) and `ade4` v1.7-22 (Dray & Dufour, 2007).

Geographic enrichment assessment of genetic groups

Collection metadata was used to derive geographic origin for each accession, with a small number from an unknown origin ($n = 180$). Geographic enrichment of each genetic group was assessed using individual ancestry proportions (Q -values) as continuous response variables. For each group, differences in Q -value distributions among geographic regions were tested using non-parametric Kruskal–Wallis rank-sum tests and, when significant, pairwise Wilcoxon rank-sum tests with a Bonferroni correction were applied to evaluate significant enrichment based on geography. To account for sample bias, we rarefied each region to 50 accessions without replacement across 500 replicates and tested for enrichment using the methodology mentioned above. After assessing broad-scale geographic enrichment, fine-scale geographic regions were assessed using the same two tests using the `kruskal.test` and `pairwise.wilcox.test` from the `stats` package in R 4.3.0 (R Core Team, 2023). Regions with fewer than 10 accessions, including the British Isles, Micronesia, Polynesia and Western Europe, were removed prior to analysis.

ACKNOWLEDGEMENTS

We thank Virginia McQueen and Erica Steadman at the Australian Grains Genebank (AGG) for assisting with tissue sampling. We thank Shyam Tallury and the United States Department of Agriculture (USDA) Genebank; Olaniyi Oyatomi and the International Institute of Tropical Agriculture (IITA) Genebank; the National Agriculture and Food Research Organization (NARO) Genebank; Rogerio Chiulele from Eduardo Mondlane University of Mozambique (EMU); the Australian Grains Genebank (AGG); Timothy Close and Maria Munoz-Amatriain for providing UCR germplasm; and Jean-Philippe Vielle-Calzada and his team at Cinvestav Langebio for help collecting and providing cowpea accessions.

This work was supported in part by the Gates Foundation (Hy-Gain INV-002955). The conclusions and opinions expressed in this work are those of the author(s) alone and shall not be attributed to the Foundation. Under the grant conditions of the Foundation, a Creative Commons Attribution 4.0 License has already been assigned to the Author Accepted Manuscript version that might arise from this submission. Please note works submitted as a preprint have not undergone a peer review process. . Open access publishing facilitated by The University of Queensland, as part of the Wiley - The University of Queensland agreement via the Council of Australasian University Librarians

AUTHOR CONTRIBUTIONS

EM, DJ, AC and AK designed the project. PM, MMW, AC, MD'A, TI, JC, PO-A, J-PV-C and SN collected and provided the plant materials. TS prepared the sequenced samples. SP, AH and SS performed data analyses. SP wrote the manuscript. EM, DJ, MD'A, MMW, SS, AH, AK, PO-A, YT, JC, TI and SN revised the manuscript.

CONFLICT OF INTEREST

The authors declare no competing interests.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available at UQ eSpace in the dataset 'CowpeaDiversityGenetic-Data' using the following link: <https://doi.org/10.48610/>

659a65b, in the [Supporting Information](#) of this article, and in the following GitHub repository: <https://github.com/SofiePearson/CowpeaDiversity>. Genomic data are currently being integrated and can be accessed through an interactive web application called AGG Pretzel (<https://agg.plantinformatix.io/>).

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article.

Appendix S1. Summary of the cowpea phylogenetic tree and geographic relationships.

Figure S1. Heatmap showing the density and distribution of 4290 SNP markers across the 11 cowpea chromosomes.

Figure S2. Principal component (PC) analysis of 10 617 accessions with each panel representing various metadata.

Figure S3. Support for the number of ancestral populations (K) from the ADMIXTURE analysis where K was tested from 1 to 40.

Figure S4. Probability support for the number of ancestral populations (K) from the STRUCTURE analysis where K was tested from 1 to 25.

Figure S5. Cross-entropy support for the number of ancestral populations (K) from the sNMF analysis where K was tested from 1 to 25.

Figure S6. Geographic distribution of the Populations at $K = 2$.

Figure S7. Geographic distribution of the Populations at $K = 3$.

Figure S8. Principal component (PC) analysis of the 9609 cowpea accessions coloured by group ($K = 9$).

Figure S9. Maximum likelihood phylogeny of 9610 cowpea accessions implemented in IQ-TREE.

Figure S10. Maximum likelihood phylogeny of 894 cowpea accessions from Clade 15.

Figure S11. Maximum likelihood phylogeny of 1056 cowpea accessions from Clade 17.

Figure S12. Maximum likelihood phylogeny of 384 cowpea accessions from Clade 20.

Figure S13. Maximum likelihood phylogeny of 733 cowpea accessions from Clade 27.

Figure S14. Maximum likelihood phylogeny of 1874 cowpea accessions from Clade 29.

Figure S15. Maximum likelihood phylogeny of 1241 cowpea accessions from Clade 30.

Figure S16. Maximum likelihood phylogeny of 712 cowpea accessions from Clade 33.

Figure S17. Maximum likelihood phylogeny of 748 cowpea accessions from Clade 34.

Table S1. Passport descriptors of 10 618 cowpea accessions.

Table S2. Percentage of improvement status category assignment for 10 617 cultivated cowpea accessions for each collection source.

Table S3. Paired potential duplicate accessions (33 252) and paired duplicate accessions (1329) identified from genetic distance (GD), genomic relatedness (GRM), whole-genome identity by descent (PI HAT, Z0, Z1 and Z2), and accession name and alias matching.

Table S4. Admixture proportion assignment of each accession to three values of K .

Table S5. Non-parametric assessment of regional geographic enrichment of group proportions for nine genetic groups.

Table S6. Non-parametric assessment of regional geographic enrichment of group proportions for nine genetic groups using rarefaction without replacement with 50 samples per region and 500 permutations.

Table S7. Tree data for the circular 9610 cowpea tree, including node, branch lengths, bootstrap support and accession order.

Table S8. Counts of private alleles within each genetic group at $K = 9$.

Table S9. Counts of private alleles within each continent.

Table S10. Diversity metrics for each continent.

Table S11. Predicted group membership of 1008 redundant cowpea accessions at three levels of K .

Table S12. Number of accessions within each clade by their population assignment at $K = 2$.

Table S13. Number of accessions within each clade by their population assignment at $K = 3$.

Table S14. Number of accessions within each clade by their group assignment at $K = 9$.

Table S15. Percentage of accessions within each clade by their geographic provenance.

Table S16. Percentage of accessions within each clade by their improvement status.

Table S17. Percentage of accessions within each clade by their collection source.

Table S18. Percentage of accessions within each clade by their cultivar group.

Table S19. Regional geographic locations for American (Northern, Central and Southern) accessions and their corresponding clade.

REFERENCES

- Abebe, B.K. & Alemayehu, M.T. (2022) A review of the nutritional use of cowpea (*Vigna unguiculata* L. Walp) for human and animal diets. *Journal of Agriculture and Food Research*, **10**, 100383. Available from: <https://doi.org/10.1016/j.jafr.2022.100383>
- Albuquerque, H.Y.G.D., Oliveira, E.J.D., Brito, A.C., Andrade, L.R.B.D., Carmo, C.D.D., Morgante, C.V. *et al.* (2019) Identification of duplicates in cassava germplasm banks based on single-nucleotide polymorphisms (SNPs). *Scientia Agricola*, **76**, 328–336. Available from: <https://doi.org/10.1590/1678-992X-2017-0389>
- Alexander, D.H., Novembre, J. & Lange, K. (2009) Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*, **19**, 1655–1664. Available from: <https://doi.org/10.1101/gr.094052.109>
- Amadeu, R.R., Garcia, A.a.F., Munoz, P.R. & Ferrao, L.F.V. (2023) AGHmatrix: genetic relationship matrices in R. *Bioinformatics*, **39**, btad445. Available from: <https://doi.org/10.1093/bioinformatics/btad445>
- Anglin, N.L., Wenzl, P., Azevedo, V., Lusty, C., Ellis, D. & Gao, D. (2025) Genotyping genebank collections: strategic approaches and considerations for optimal collection management. *Plants*, **14**, 252. Available from: <https://doi.org/10.3390/plants14020252>
- Asdal, Å. & Guarino, L. (2018) The Svalbard global seed vault: 10 years—1 million samples. *Biopreservation and Biobanking*, **16**, 391–392. Available from: <https://doi.org/10.1089/bio.2018.0025>
- Bjornson, S., Verbruggen, H., Upham, N.S. & Steenwyk, J.L. (2024) Reticulate evolution: detection and utility in the phylogenomics era. *Molecular Phylogenetics and Evolution*, **201**, 108197. Available from: <https://doi.org/10.1016/j.ympev.2024.108197>
- Bolarinwa, K.A., Ogunkanmi, L.A., Ogundipe, O.T., Agboola, O.O. & Amusa, O.D. (2022) An investigation of cowpea production constraints and preferences among small holder farmers in Nigeria. *GeoJournal*, **87**, 2993–3005. Available from: <https://doi.org/10.1007/s10708-021-10405-6>
- Boukar, O., Abberton, M., Oyatomi, O., Togola, A., Tripathi, L. & Fatokun, C. (2020) Introgression breeding in cowpea [*Vigna unguiculata* (L.) Walp.]. *Frontiers in Plant Science*, **11**, 567425. Available from: <https://doi.org/10.3389/fpls.2020.567425>

- Boukar, O., Belko, N., Chamarthi, S., Togola, A., Batiemo, J., Owusu, E. et al. (2019) Cowpea (*Vigna unguiculata*): genetics, genomics and breeding. *Plant Breeding*, **138**, 415–424. Available from: <https://doi.org/10.1111/pbr.12589>
- Boukar, O., Fatokun, C.A., Huynh, B.-L., Roberts, P.A. & Close, T.J. (2016) Genomic tools in cowpea breeding programs: status and perspectives. *Frontiers in Plant Science*, **7**, 757. Available from: <https://doi.org/10.3389/fpls.2016.00757>
- Browning, B.I. & Browning, S.R. (2016) Genotype imputation with millions of reference samples. *The American Journal of Human Genetics*, **98**, 116–126. Available from: <https://doi.org/10.1016/j.ajhg.2015.11.020>
- Carvalho, M., Muñoz-Amatriain, M., Castro, I., Lino-Neto, T., Matos, M., Egea-Cortines, M. et al. (2017) Genetic diversity and structure of Iberian Peninsula cowpeas compared to world-wide cowpea accessions using high density SNP markers. *BMC Genomics*, **18**, 891. Available from: <https://doi.org/10.1186/s12864-017-4295-0>
- Chen, H., Chen, H., Hu, L., Wang, L., Wang, S., Wang, M.L. et al. (2017) Genetic diversity and a population structure analysis of accessions in the Chinese cowpea [*Vigna unguiculata* (L.) Walp.] germplasm collection. *The Crop Journal*, **5**, 363–372. Available from: <https://doi.org/10.1016/j.cj.2017.04.002>
- Chipeta, M.M., Kafwambira, J. & Yohane, E. (2025) Cowpea genetic diversity, population structure and genome-wide association studies in Malawi: insights for breeding programs. *Frontiers in Plant Science*, **15**, 1461631. Available from: <https://doi.org/10.3389/fpls.2024.1461631>
- Cruz, V.M.V., Kilian, A. & Dierig, D.A. (2013) Development of DArT marker platforms and genetic diversity assessment of the U.S. collection of the new oilseed crop *Lesquerella* and related species. *PLoS One*, **8**, e64062. Available from: <https://doi.org/10.1371/journal.pone.0064062>
- D'andrea, A.C., Kahlheber, S., Logan, A.L. & Watson, D.J. (2007) Early domesticated cowpea (*Vigna unguiculata*) from Central Ghana. *Antiquity*, **81**, 686–698. Available from: <https://doi.org/10.1017/S0003598X00095661>
- Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A. et al. (2011) The variant call format and VCFtools. *Bioinformatics*, **27**, 2156–2158. Available from: <https://doi.org/10.1093/bioinformatics/btr330>
- Danecek, P., Bonfield, J.K., Liddle, J., Marshall, J., Ohan, V., Pollard, M.O. et al. (2021) Twelve years of SAMtools and BCFtools. *GigaScience*, **10**, giab008. Available from: <https://doi.org/10.1093/gigascience/giab008>
- Degnan, J.H. & Rosenberg, N.A. (2009) Gene tree discordance, phylogenetic inference and the multispecies coalescent. *Trends in Ecology & Evolution*, **24**, 332–340. Available from: <https://doi.org/10.1016/j.tree.2009.01.009>
- Desalegne, B.A., Mohammed, S., Dagne, K. & Timko, M.P. (2016) Assessment of genetic diversity in Ethiopian cowpea [*Vigna unguiculata* (L.) Walp.] germplasm using simple sequence repeat markers. *Plant Molecular Biology Reporter*, **34**, 978–992. Available from: <https://doi.org/10.1007/s11105-016-0979-x>
- Dixit, G.P. & Satheesh Naik, S.J. (2025) Global research status of grain legumes. In: Prasad Dixit, G., Dikshit, H.K., Mishra, G.P. & Aski, M.S. (Eds.) *Fundamentals of legume breeding: a text for students and practitioners*. Singapore: Springer Nature Singapore.
- Doyle, J.J. & Doyle, J.L. (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin*, **19**, 11–15.
- Dray, S. & Dufour, A.-B. (2007) The ade4 package: implementing the duality diagram for ecologists. *Journal of Statistical Software*, **22**, 1–20. Available from: <https://doi.org/10.18637/jss.v022.i04>
- Egbadzor, K.F., Ofori, K., Yeboah, M., Aboagye, L.M., Opoku-Agyeman, M.O., Danquah, E.Y. et al. (2014) Diversity in 113 cowpea [*Vigna unguiculata* (L.) Walp.] accessions assessed with 458 SNP markers. *Springerplus*, **3**, 541. Available from: <https://doi.org/10.1186/2193-1801-3-541>
- Evanno, G., Regnaut, S. & Goudet, J. (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology*, **14**, 2611–2620. Available from: <https://doi.org/10.1111/j.1365-294x.2005.02553.x>
- Fang, J., Chao, C.-C.T., Roberts, P.A. & Ehlers, J.D. (2007) Genetic diversity of cowpea [*Vigna unguiculata* (L.) Walp.] in four West African and USA breeding programs as determined by AFLP analysis. *Genetic Resources and Crop Evolution*, **54**, 1197–1209. Available from: <https://doi.org/10.1007/s10722-006-9101-9>
- FAO. (2025) FAOSTAT statistical database [Online]. Available from: <https://www.fao.org/faostat/en/#data/QCL> [Accessed January 6, 2026]
- Fatokun, C., Girma, G., Abberton, M., Gedil, M., Unachukwu, N., Oyatomi, O. et al. (2018) Genetic diversity and population structure of a mini-core subset from the world cowpea (*Vigna unguiculata* (L.) Walp.) germplasm collection. *Scientific Reports*, **8**, 16035. Available from: <https://doi.org/10.1038/s41598-018-34555-9>
- Fatokun, C.A. & Ng, Q. (2007) Outcrossing in cowpea. *Journal of Food, Agriculture & Environment*, **5**, 334–338.
- Fiscus, C.J., Herniter, I.A., Tchamba, M., Paliwal, R., Muñoz-Amatriain, M., Roberts, P.A. et al. (2024) The pattern of genetic variability in a core collection of 2,021 cowpea accessions. *G3 (Bethesda, Md.)*, **14**, jkae071. Available from: <https://doi.org/10.1093/g3journal/jkae071>
- Francis, R.M. (2017) Pophelper: an R package and web app to analyse and visualize population structure. *Molecular Ecology Resources*, **17**, 27–32. Available from: <https://doi.org/10.1111/1755-0998.12509>
- Frichot, E. & François, O. (2015) LEA: an R package for landscape and ecological association studies. *Methods in Ecology and Evolution*, **6**, 925–929. Available from: <https://doi.org/10.1111/2041-210X.12382>
- Frichot, E., Mathieu, F., Trouillon, T., Bouchard, G. & François, O. (2014) Fast and efficient estimation of individual ancestry coefficients. *Genetics*, **196**, 973–983. Available from: <https://doi.org/10.1534/genetics.113.160572>
- Fuller, D.Q. (2003) *African crops in prehistoric South Asia: a critical review*.
- Gbedevi, K.M., Boukar, O., Ishikawa, H., Abe, A., Ongom, P.O., Unachukwu, N. et al. (2021) Genetic diversity and population structure of cowpea [*Vigna unguiculata* (L.) Walp.] germplasm collected from Togo based on DArT markers. *Genes (Basel)*, **12**, 1451. Available from: <https://doi.org/10.3390/genes12091451>
- Goldstein, S.T., Mueller, N.G., Janzen, A., Ogola, C., Dal Martello, R., Fernandes, R. et al. (2024) Early agriculture and crop transitions at Kakapel Rockshelter in the Lake Victoria region of eastern Africa. *Proceedings of the Royal Society B: Biological Sciences*, **291**, 20232747. Available from: <https://doi.org/10.1098/rspb.2023.2747>
- Goudet, J. (2005) Hierfstat, a package for r to compute and test hierarchical F-statistics. *Molecular Ecology Notes*, **5**, 184–186. Available from: <https://doi.org/10.1111/j.1471-8286.2004.00828.x>
- Gruber, K. (2017) Agrobiodiversity: the living library. *Nature*, **544**, S8–S10. Available from: <https://doi.org/10.1038/544S8a>
- Gu, D., Andreev, K. & Dupre, M.E. (2021) Major trends in population growth around the world. *China CDC Weekly*, **3**, 604–613. Available from: <https://doi.org/10.46234/ccdcw2021.160>
- Gumede, M.T., Gerrano, A.S., Amelework, A.B. & Modi, A.T. (2022) Analysis of genetic diversity and population structure of cowpea (*Vigna unguiculata* (L.) Walp.) genotypes using single nucleotide polymorphism markers. *Plants*, **11**, 3480. Available from: <https://doi.org/10.3390/plants11243480>
- Hay, F.R. & Probert, R.J. (2013) Advances in seed conservation of wild plant species: a review of recent research. *Conservation Physiology*, **1**, cot030. Available from: <https://doi.org/10.1093/conphys/cot030>
- Hemstrom, W. & Jones, M. (2023) snpR: user friendly population genomics for SNP data sets with categorical metadata. *Molecular Ecology Resources*, **23**, 962–973. Available from: <https://doi.org/10.1111/1755-0998.13721>
- Herniter, I.A., Muñoz-Amatriain, M. & Close, T.J. (2020) Genetic, textual, and archeological evidence of the historical global spread of cowpea (*Vigna unguiculata* [L.] Walp.). *Legume Science*, **2**, e57. Available from: <https://doi.org/10.1002/leg3.57>
- Hibbins, M.S. & Hahn, M.W. (2021) Phylogenomic approaches to detecting and characterizing introgression. *Genetics*, **220**, iyab173. Available from: <https://doi.org/10.1093/genetics/iyab173>
- Hoang, D.T., Chernomor, O., Von Haeseler, A., Minh, B.Q. & Vinh, L.S. (2018) UFBoot2: improving the ultrafast bootstrap approximation. *Molecular Biology and Evolution*, **35**, 518–522. Available from: <https://doi.org/10.1093/molbev/msx281>
- Hohenlohe, P.A., Bassham, S., Etter, P.D., Stiffler, N., Johnson, E.A. & Cresko, W.A. (2010) Population genomics of parallel adaptation in three-spine stickleback using sequenced RAD tags. *PLoS Genetics*, **6**, e1000862. Available from: <https://doi.org/10.1371/journal.pgen.1000862>
- Hurlbert, S.H. (1971) The nonconcept of species diversity: a critique and alternative parameters. *Ecology*, **52**, 577–586. Available from: <https://doi.org/10.2307/1934145>
- Huynh, B.-L., Close, T.J., Roberts, P.A., Hu, Z., Wanamaker, S., Lucas, M.R. et al. (2013) Gene pools and the genetic architecture of domesticated cowpea. *The Plant Genome*, **6**, plantgenome2013.03.0005. Available from: <https://doi.org/10.3835/plantgenome2013.03.0005>

- Janes, J.K., Miller, J.M., Dupuis, J.R., Malenfant, R.M., Gorrell, J.C., Cullingham, C.I. *et al.* (2017) The $K=2$ conundrum. *Molecular Ecology*, **26**, 3594–3602. Available from: <https://doi.org/10.1111/mec.14187>
- Jayathilake, C., Visvanathan, R., Deen, A., Bangamuwage, R., Jayawardana, B.C., Nammi, S. *et al.* (2018) Cowpea: an overview on its nutritional facts and health benefits. *Journal of the Science of Food and Agriculture*, **98**, 4793–4806. Available from: <https://doi.org/10.1002/jsfa.9074>
- Kalyaanamoorthy, S., Minh, B.Q., Wong, T.K.F., Von Haeseler, A. & Jermini, L.S. (2017) ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature Methods*, **14**, 587–589. Available from: <https://doi.org/10.1038/nmeth.4285>
- Kamvar, Z.N., Brooks, J.C. & Grünwald, N.J. (2015) Novel R tools for analysis of genome-wide population genetic data with emphasis on clonality. *Frontiers in Genetics*, **6**, 208. Available from: <https://doi.org/10.3389/fgene.2015.00208>
- Kamvar, Z.N., Tabima, J.F. & Grünwald, N.J. (2014) Poppr: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ*, **2**, e281. Available from: <https://doi.org/10.7717/peerj.281>
- Ketema, S., Tesfaye, B., Keneni, G., Amsalu Fenta, B., Assefa, E., Greliche, N. *et al.* (2020) DArTSeq SNP-based markers revealed high genetic diversity and structured population in Ethiopian cowpea [*Vigna unguiculata* (L.) Walp.] germplasm. *PLoS One*, **15**, e0239122. Available from: <https://doi.org/10.1371/journal.pone.0239122>
- Khazaei, H., Street, K., Bari, A., Mackay, M. & Stoddard, F.L. (2013) The FIGS (focused identification of germplasm strategy) approach identifies traits related to drought adaptation in *Vicia faba* genetic resources. *PLoS One*, **8**, e63107. Available from: <https://doi.org/10.1371/journal.pone.0063107>
- Kilian, A., Wenzl, P., Huttner, E., Carling, J., Xia, L., Blois, H. *et al.* (2012) Diversity arrays technology: a generic genome profiling technology on open platforms. In: Pompanon, F. & Bonin, A. (Eds.) *Data production and analysis in population genomics: methods and protocols*. Totowa, NJ: Humana Press.
- Kim, D.-K., Ochar, K., Iwar, K., Ha, B.-K. & Kim, S.-H. (2025) Cowpea (*Vigna unguiculata* L.) production, genetic resources and strategic breeding priorities for sustainable food security: a review. *Frontiers in Plant Science*, **16**, 1562142. Available from: <https://doi.org/10.3389/fpls.2025.1562142>
- Kussie, B., Tadele, Y. & Asresie, A. (2024) Effect of maize (*Zea mays* L.) and cowpea (*Vigna unguiculata* L.) intercropping on agronomic performance, yield and nutritional values of maize and cowpea under supplementary irrigation. *Heliyon*, **10**, e39817. Available from: <https://doi.org/10.1016/j.heliyon.2024.e39817>
- Li, H. & Durbin, R. (2009) Fast and accurate short read alignment with burrows-wheeler transform. *Bioinformatics*, **25**, 1754–1760. Available from: <https://doi.org/10.1093/bioinformatics/btp324>
- Liang, Q., Muñoz-Amatriáin, M., Shu, S., Lo, S., Wu, X., Carlson, J.W. *et al.* (2024) A view of the pan-genome of domesticated cowpea (*Vigna unguiculata* [L.] Walp.). *The Plant Genome*, **17**, e20319. Available from: <https://doi.org/10.1002/tpg2.20319>
- Liu, L., Xi, Z. & Davis, C.C. (2014) Coalescent methods are robust to the simultaneous effects of long branches and incomplete lineage sorting. *Molecular Biology and Evolution*, **32**, 791–805. Available from: <https://doi.org/10.1093/molbev/msu331>
- Lonardi, S., Muñoz-Amatriáin, M., Liang, Q., Shu, S., Wanamaker, S.I., Lo, S. *et al.* (2019) The genome of cowpea (*Vigna unguiculata* [L.] Walp.). *The Plant Journal*, **98**, 767–782. Available from: <https://doi.org/10.1111/tpj.14349>
- Macharia, M.W., Caproni, L., Solemanegy, M., Takele, R., Chiulele, R.M., Munisse, P. *et al.* (2025) Cowpea (*Vigna unguiculata* L. Walp.) landraces in Mozambique and neighbouring southern African countries harbour genetic loci with potential for climate adaptation. *Communications Biology*, **8**, 834. Available from: <https://doi.org/10.1038/s42003-025-08227-0>
- Mahalakshmi, V., Ng, Q., Lawson, M. & Ortiz, R. (2007) Cowpea [*Vigna unguiculata* (L.) Walp.] core collection defined by geographical, agronomical and botanical descriptors. *Plant Genetic Resources*, **5**, 113–119. Available from: <https://doi.org/10.1017/S1479262107837166>
- Mallet, J. (2005) Hybridization as an invasion of the genome. *Trends in Ecology & Evolution*, **20**, 229–237. Available from: <https://doi.org/10.1016/j.tree.2005.02.010>
- Mascher, M., Schreiber, M., Scholz, U., Graner, A., Reif, J.C. & Stein, N. (2019) Genebank genomics bridges the gap between the conservation of crop diversity and plant breeding. *Nature Genetics*, **51**, 1076–1081. Available from: <https://doi.org/10.1038/s41588-019-0443-6>
- Mccouch, S., McNally, K., Wang, W. & Hamilton, R. (2012) Genomics of gene Banks: a case study in rice. *American Journal of Botany*, **99**, 407–423. Available from: <https://doi.org/10.3732/ajb.1100385>
- Milner, S.G., Jost, M., Taketa, S., Mazón, E.R., Himmelbach, A., Oppermann, M. *et al.* (2019) Genebank genomics highlights the diversity of a global barley collection. *Nature Genetics*, **51**, 319–326. Available from: <https://doi.org/10.1038/s41588-018-0266-x>
- Minh, B.Q., Schmidt, H.A., Chernomor, O., Schrempf, D., Woodhams, M.D., Von Haeseler, A. *et al.* (2020) IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Molecular Biology and Evolution*, **37**, 1530–1534. Available from: <https://doi.org/10.1093/molbev/msaa015>
- Muñoz-Amatriáin, M., Lo, S., Herniter, I.A., Boukar, O., Fatokun, C., Carvalho, M. *et al.* (2021) The UCR Minicore: a resource for cowpea research and breeding. *Legume Science*, **3**, e95. Available from: <https://doi.org/10.1002/leg3.95>
- Nkhoma, N., Shimelis, H., Laing, M.D., Shayanowako, A. & Mathew, I. (2020) Assessing the genetic diversity of cowpea [*Vigna unguiculata* (L.) Walp.] germplasm collections using phenotypic traits and SNP markers. *BMC Genetics*, **21**, 110. Available from: <https://doi.org/10.1186/s12863-020-00914-7>
- Ofem, N.P., Kamal, N.M., Pearson, S., Shatte, T., Jordan, D., Mace, E. *et al.* (2025) Genetic diversity and structure of Asian cowpea germplasm. *Scientific Reports*, **15**, 27909. Available from: <https://doi.org/10.1038/s41598-025-13511-4>
- Ongom, P.O., Fatokun, C., Togola, A., Mohammed, S.B., Ishaya, D.J., Bala, G. *et al.* (2023) Exploiting the genetic potential of cowpea in an intercropping complex. *Agronomy*, **13**, 1594. Available from: <https://doi.org/10.3390/agronomy13061594>
- Padulosi, S. & Ng, N. (1997) Origin, taxonomy, and morphology of *Vigna unguiculata* (L.) Walp. In: Singh, B.B., Mohan Raj, D., Dashiell, K. & Jackai, L.E.N. (Eds.) *Advances in cowpea research*. Ibadan, Nigeria: IITA.
- Pasquet, R.S. (2000) Allozyme diversity of cultivated cowpea *Vigna unguiculata* (L.) Walp. *Theoretical and Applied Genetics*, **101**, 211–219. Available from: <https://doi.org/10.1007/s001220051471>
- Philippe, H., Brinkmann, H., Lavrov, D.V., Littlewood, D.T.J., Manuel, M., Wörheide, G. *et al.* (2011) Resolving difficult phylogenetic questions: why more sequences are not enough. *PLoS Biology*, **9**, e1000602. Available from: <https://doi.org/10.1371/journal.pbio.1000602>
- Pritchard, J.K., Stephens, M. & Donnelly, P. (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945–959. Available from: <https://doi.org/10.1093/genetics/155.2.945>
- Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.a., Bender, D. *et al.* (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics*, **81**, 559–575. Available from: <https://doi.org/10.1086/519795>
- R Core Team. (2023) *R: a language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Raman, H., Raman, R., Kilian, A., Detering, F., Carling, J., Coombes, N. *et al.* (2014) Genome-wide delineation of natural variation for pod shatter resistance in *Brassica napus*. *PLoS One*, **9**, e101673. Available from: <https://doi.org/10.1371/journal.pone.0101673>
- Ray, D.K., Mueller, N.D., West, P.C. & Foley, J.A. (2013) Yield trends are insufficient to double global crop production by 2050. *PLoS One*, **8**, e66428. Available from: <https://doi.org/10.1371/journal.pone.0066428>
- Sansaloni, C., Franco, J., Santos, B., Percival-Alwyn, L., Singh, S., Petroli, C. *et al.* (2020) Diversity analysis of 80,000 wheat accessions reveals consequences and opportunities of selection footprints. *Nature Communications*, **11**, 4572. Available from: <https://doi.org/10.1038/s41467-020-18404-w>
- Sansaloni, C., Petroli, C., Jaccoud, D., Carling, J., Grattapaglia, D. *et al.* (2011) Diversity arrays technology (DArT) and next-generation sequencing combined: genome-wide, high throughput, highly informative genotyping for molecular breeding of eucalyptus. *BMC Proceedings*, **5**, P54. Available from: <https://doi.org/10.1186/1753-6561-5-S7-P54>
- Schrempf, D., Minh, B.Q., De Maio, N., Von Haeseler, A. & Kosiol, C. (2016) Reversible polymorphism-aware phylogenetic models and their application to tree inference. *Journal of Theoretical Biology*, **407**, 362–370. Available from: <https://doi.org/10.1016/j.jtbi.2016.07.042>

- Schrempf, D., Minh, B.Q., Von Haeseler, A. & Kosiol, C. (2019) Polymorphism-aware species trees with advanced mutation models, bootstrap, and rate heterogeneity. *Molecular Biology and Evolution*, **36**, 1294–1301. Available from: <https://doi.org/10.1093/molbev/msz043>
- Schulthess, A.W., Kale, S.M., Liu, F., Zhao, Y., Philipp, N., Rembe, M. *et al.* (2022) Genomics-informed prebreeding unlocks the diversity in genebanks for wheat improvement. *Nature Genetics*, **54**, 1544–1552. Available from: <https://doi.org/10.1038/s41588-022-01189-7>
- Seo, E., Kim, K., Jun, T.-H., Choi, J., Kim, S.-H., Muñoz-Amatriain, M. *et al.* (2020) Population structure and genetic diversity in Korean cowpea germplasm based on SNP markers. *Plants*, **9**, 1190. Available from: <https://doi.org/10.3390/plants9091190>
- Singh, B.B., Chambliss, O.L. & Sharma, B. (1997) Recent advances in cowpea breeding. In: Singh, B.B., Mohan Raji, D.R. & Dashiel, K.E. (Eds.) *Advances in cowpea research*. Ibadan, Nigeria: IITA.
- Singh, S. & Allen, D.J. (1979) *Cowpea pests and diseases*. Ibadan, Nigeria: International Institute of Tropical Agriculture.
- Smith, S., Bubeck, D., Nelson, B., Stanek, J. & Gerke, J. (2015) Genetic diversity and modern plant breeding. In: Ahuja, M.R. & Jain, S.M. (Eds.) *Genetic diversity and erosion in plants: indicators and prevention*. Cham: Springer International Publishing.
- Sodedji, F.a.K., Agbahoungba, S., Agoyi, E.E., Kafoutchoni, M.K., Choi, J., Nguetta, S.-P.A. *et al.* (2021) Diversity, population structure, and linkage disequilibrium among cowpea accessions. *The Plant Genome*, **14**, e20113. Available from: <https://doi.org/10.1002/tpg2.20113>
- Song, Q., Hyten, D.L., Jia, G., Quigley, C.V., Fickus, E.W., Nelson, R.L. *et al.* (2015) Fingerprinting soybean germplasm and its utility in genomic research. *G3: Genes, Genomes, Genetics*, **5**, 1999–2006. Available from: <https://doi.org/10.1534/g3.115.019000>
- South, A. (2011) Rworldmap: a new R package for mapping global data. *The R Journal*, **3**, 35–43. Available from: <https://doi.org/10.32614/RJ-2011-006>
- Steele, W. (1976) Cowpeas. *Vigna unguiculata*. In: Simmonds, N. (Ed.) *Evolution of crop plants*. London: Longman.
- Sunitha, N.C., Prathibha, M.D., Thribhuvan, R., Lokeshkumar, B.M., Basavaraj, P.S., Lohithaswa, H.C. *et al.* (2024) Focused identification of germplasm strategy (FIGS): a strategic approach for trait-enhanced prebreeding. *Genetic Resources and Crop Evolution*, **71**, 1–16. Available from: <https://doi.org/10.1007/s10722-023-01669-7>
- Tadele, Z. (2019) Orphan crops: their importance and the urgency of improvement. *Planta*, **250**, 677–694. Available from: <https://doi.org/10.1007/s00425-019-03210-6>
- Tanksley, S.D. & McCouch, S.R. (1997) Seed Banks and molecular maps: unlocking genetic potential from the wild. *Science*, **277**, 1063–1066. Available from: <https://doi.org/10.1126/science.277.5329.1063>
- Tilman, D., Balzer, C., Hill, J. & Befort, B.L. (2011) Global food demand and the sustainable intensification of agriculture. *Proceedings of the National Academy of Sciences*, **108**, 20260–20264. Available from: <https://doi.org/10.1073/pnas.1116437108>
- Van Dijk, M., Morley, T., Rau, M.L. & Saghai, Y. (2021) A meta-analysis of projected global food demand and population at risk of hunger for the period 2010–2050. *Nature Food*, **2**, 494–501. Available from: <https://doi.org/10.1038/s43016-021-00322-9>
- Weir, B.S. & Cockerham, C.C. (1984) Estimating F-statistics for the analysis of population structure. *Evolution*, **38**, 1358–1370. Available from: <https://doi.org/10.2307/2408641>
- Whitfield, J.B. & Lockhart, P.J. (2007) Deciphering ancient rapid radiations. *Trends in Ecology & Evolution*, **22**, 258–265. Available from: <https://doi.org/10.1016/j.tree.2007.01.012>
- Wickham, H. (2016) *ggplot2: elegant graphics for data analysis*. New York: Springer-Verlag.
- Xiong, H., Shi, A., Mou, B., Qin, J., Motes, D., Lu, W. *et al.* (2016) Genetic diversity and population structure of cowpea (*Vigna unguiculata* L. Walp). *PLoS One*, **11**, e0160941. Available from: <https://doi.org/10.1371/journal.pone.0160941>
- Xu, S., Dai, Z., Guo, P., Fu, X., Liu, S., Zhou, L. *et al.* (2021) ggtreeExtra: compact visualization of richly annotated phylogenetic data. *Molecular Biology and Evolution*, **38**, 4039–4042. Available from: <https://doi.org/10.1093/molbev/msab166>
- Yang, M., Guarín, J.R., Freduah, B.S., Wesley, G.O., Maccarthy, D.S., Narh, S. *et al.* (2025) Climate-crop models to support opportunity crop adaptation in Africa. *Nature Communications*, **16**, 11186. Available from: <https://doi.org/10.1038/s41467-025-66180-2>
- Ye, C.-Y. & Fan, L. (2021) Orphan crops and their wild relatives in the genomic era. *Molecular Plant*, **14**, 27–39. Available from: <https://doi.org/10.1016/j.molp.2020.12.013>
- Yu, G., Smith, D.K., Zhu, H., Guan, Y. & Lam, T.T.-Y. (2017) Ggtree: an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods in Ecology and Evolution*, **8**, 28–36. Available from: <https://doi.org/10.1111/2041-210X.12628>