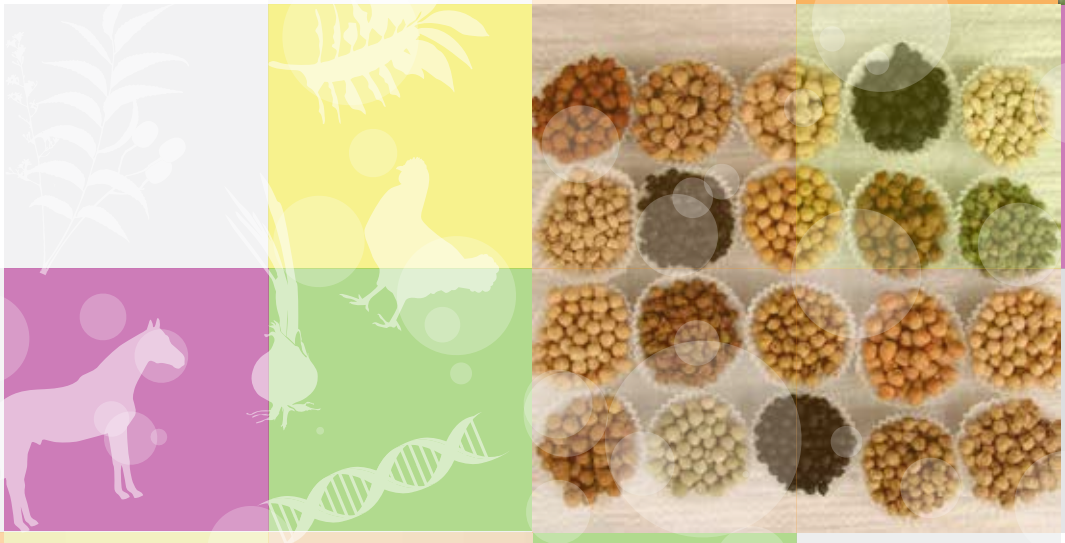
Genetic Resources



Genetic Resources (2026)
Vol. 7, Issue 13
DOI: 10.46265/genresj.2026.13
www.genresj.org
ISSN: 2708-3764

Focus and Scope of *Genetic Resources*

Genetic Resources is an open access journal disseminating global knowledge and tools used by the community of practitioners of plant and animal genetic resources involved in monitoring, collecting, maintaining, conserving, characterizing and using genetic resources for food, agriculture and forestry. ***Genetic Resources*** publishes original research, methods, strategies, guidelines, case studies and reviews as well as opinion and other papers on a variety of topics of interest on the present and future use of genetic resources. These may include the acquisition, documentation, conservation, management, assessment, characterization and evaluation of genetic resources and their link to broader biodiversity, socioeconomic practices, policy guidelines or similar, serving stakeholders within and across sectors. Occasionally, ***Genetic Resources*** publishes special issues with a focus on selected topics of interest for the genetic resources community. The journal has a focus on the European region and also welcomes contributions of wider interest from all world regions.

Cover photos (from left to right):
Crotalaria accessions from Embrapa genebank, Brazil. Credit: Machado Squarisi *et al*; Chickpea accessions from Ukraine. Credit: Vus *et al*; Nordic horses. Credit: Smogeli *et al*.

This journal is supported by the European Cooperative Programme for Plant Genetic Resources (ECPGR) and the European Regional Focal Point for Animal Genetic Resources (ERFP).

This journal was conceived as part of the GenRes Bridge project. This project received seed funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 817580.

Editorial Office

ECPGR Secretariat
c/o Alliance of Bioversity International and CIAT
Via di San Domenico 1
000153 Rome, Italy

Submissions to www.genresj.org

Editorial Board

Genetic Resources

Genetic Resources (2025), 7 (13)
DOI: 10.46265/genresj.2026.13
www.genresj.org
ISSN: 2708-3764

TABLE OF CONTENTS

Original Articles

Reviews

Other articles

Genetic Resources

OTHER ARTICLES

Genetic Resources (2026), 7 (13), 1–11
DOI: 10.46265/genresj.HLES2857
https://www.genresj.org
ISSN: 2708-3764

Current status of the diversity and conservation of genetic resources of *Capsicum* spp. (chilli pepper) in Oaxaca, Mexico

Yeimy Clemencia Ramirez-Rodas*, Luis Yobani Gayosso-Rosales and Ulises Santiago-López

Instituto Nacional de Investigaciones Forestales, Agrícolas y Pecuarias. C.E. Valles Centrales, C. Melchor Ocampo 7, Sto. Domingo Barrio Bajo, Etla, Oaxaca, México. C.P. 68200

Abstract: Oaxaca, Mexico, has a wide diversity of genetic resources of *Capsicum* spp. (chilli pepper), with at least 25 of the 90 chilli pepper types reported for Mexico. However, some have stopped being cultivated, making it necessary to conduct an updated diagnosis to implement conservation and sustainable use practices. The objective of this work was to describe the current situation of chilli pepper crop biodiversity in the state of Oaxaca, Mexico, and identify how they are conserved. The study identified that the most important chilli peppers for production or cultural uses in Oaxaca are: Agua, Costeño, Soledad, Tabaquero, Taviche, and Huacle, as well as less utilized varieties, including domesticated, semi-wild or wild varieties that are part of the state's cultural identity and diversity. Currently, seed collections of native chilli peppers are conserved *ex situ* in three seedbanks in Mexico, and they are also cultivated (*in situ*) by farmers, often in backyards, or left uncultivated (semi-wild or wild).

Keywords: Agrobiodiversity, native plant, conservation of chilli pepper

Citation: Ramirez-Rodas, Y. C., Gayosso-Rosales, L. Y. and Santiago-López, U. (2026) "Current status of the diversity and conservation of genetic resources of *Capsicum* spp. (chilli pepper) in Oaxaca, Mexico", *Genetic Resources,* 7(13), pp. 1–11. doi: 10.46265/genresj.HLES2857.

Introduction

The genus *Capsicum* (chilli pepper) belongs to the Solanaceae family. This genus includes five domesticated species: *Capsicum chinense* Jacq., *Capsicum frutescens* L., *Capsicum baccatum* L., *Capsicum pubescens* Ruiz & Pav., and *Capsicum annuum* L., the latter encompassing most of the chilli varieties in Mexico and being the most agriculturally and economically important (Pickersgill, 2016). The use and cultivation of *C. annuum* have a long cultural history in Mexico (Pérez-Martínez *et al*, 2022). In every culture within the multiethnic territory, chilli is a central ingredient in daily cuisine, forming part of national identity (Ruiz-Núñez *et al*, 2018). Around 90 types of chilli are known in Mexico (Aguilar-Meléndez and Lira-Noriega, 2018), with the greatest diversity found in the southern states, especially Oaxaca, which harbours at least 25 types (Aguilar-Rincón *et al*, 2010). Despite their agricultural, economic, cultural and gastronomic significance, these plant genetic resources are threatened by genetic erosion, and some types have ceased to be cultivated (Rodríguez-Campos, 2018). In fact, some farmers in Oaxaca report that certain chilli types are no longer grown due to pests, diseases, replacement by commercial varieties and water scarcity, among other factors (Aguilar-Rincón *et al*, 2010). In addition, local species are often limited by the agroecological characteristics of their environment and generally remain within their natural distribution areas (Guzmán-Mendoza *et al*, 2022). In this regard, genetic resources, including plant species, are of great importance in food production and food security. Food security, from an agricultural perspective, refers to the conservation and use of biological diversity through the capacity to produce more sustainable food. This includes the use of genetic resources as raw material for genetic improvement, which involves the selection of local varieties or the development of new varieties that are more tolerant

* Corresponding authors: Yeimy Clemencia Ramirez-Rodas (ramirez.yeimy@inifap.gob.mx)

Received: 10.06.2025 | Accepted: 16.10.2025 | Published online: 15.01.2026

to drought, pests, diseases or poor soils; the selection of plant varieties with higher yield and nutritional quality; and local or improved varieties that require fewer fertilizers or pesticides (FAO, 2012). The objective of this work was to describe the current situation of chilli pepper biodiversity in the state of Oaxaca, Mexico, and to identify how it is conserved to gain a perspective for implementing exploration, conservation and use of chilli germplasm.

Chilli pepper diversity in the state of Oaxaca

Oaxaca is the Mexican state with the greatest diversity of chilli peppers, including *C. annuum* var. *annuum* (domesticated and semi-domesticated): Mirador, Nanche, Soltero, Güiña Shirundu, Mirasol, Tusta, Güiña Shuladi, Solterito, Loco, Pasilla Oaxaca (Pasilla Mixe), Costeño, Escuchito, Achilito, Huacle, Tabaquero (Chiltepe), Coxle, de Monte, Soledad, Taviche, de Onza, and de Agua; *C. annuum* var. *acuminatum* Fingerth: chile Chocolate; *C. annuum* var. *glabriusculum* (Dunal) Heiser & Pickersgill: chile Bolita and Piquín (wild relatives of domesticated varieties); and *C. pubescens*: Manzano (López-López and Castro-García, 2006; Aguilar-Rincón *et al*, 2010; Castellón-Martínez *et al*, 2012; Sclavo-Castillo *et al*, 2024). Of these, at least 44% have been identified as native to Oaxaca: Tusta, Tabaquero, Taviche, Solterito, Piquín, Nanche, Costeño, Bolita, Huacle, de Agua, and Pasilla Mixe (Vera-Guzmán *et al*, 2011; Castellón-Martínez *et al*, 2012; Martínez-Martínez *et al*, 2014; Sanjuan-Martínez, 2020). Despite their cultural, social, economic and culinary importance, some of these types are better known than others. For instance, the Statistical Yearbook of Agricultural Production from the SIAP platform (SIAP, 2023) only provides data for the de Agua, Costeño, Soledad, and Tabaquero chillis, in addition to commercial varieties such as Ancho, Habanero, Pasilla and Serrano. However, semi-domesticated types, generally cultivated in home gardens, and wild types of *C. annuum* are not included.

Moreover, the presence of the wild species *Capsicum rhomboideum* (Dunal) Kuntze [synonym *C. ciliatum* (Kunth) Kuntze] has been documented in the state of Oaxaca, mainly distributed in sub-humid zones and to a lesser extent in semi-arid and humid areas (UNAM, 2007; Aguilar-Meléndez and Lira-Noriega, 2018). These represent an important part of the genetic diversity and germplasm variability, and are deeply linked to Oaxaca's cultural heritage. Each region in the state is associated with specific types of chilli peppers, some of which are presented in detail below.

Most relevant chilli peppers based on their production and utilization

Chile de Agua

It is a native variety from the Central Valleys and is primarily cultivated in the municipalities of San Sebastián Abasolo, Heroica Ciudad de Ejutla de Crespo, San Bernardo Mixtepec, San Pablo Huixtepec, and San Jerónimo Tlacochahuaya, which together produce about 50% of the total output (Figure 1; SIAP, 2023). Geographically, it is located at 16° 59′ N and 96° 35′ W, between 1,400 and

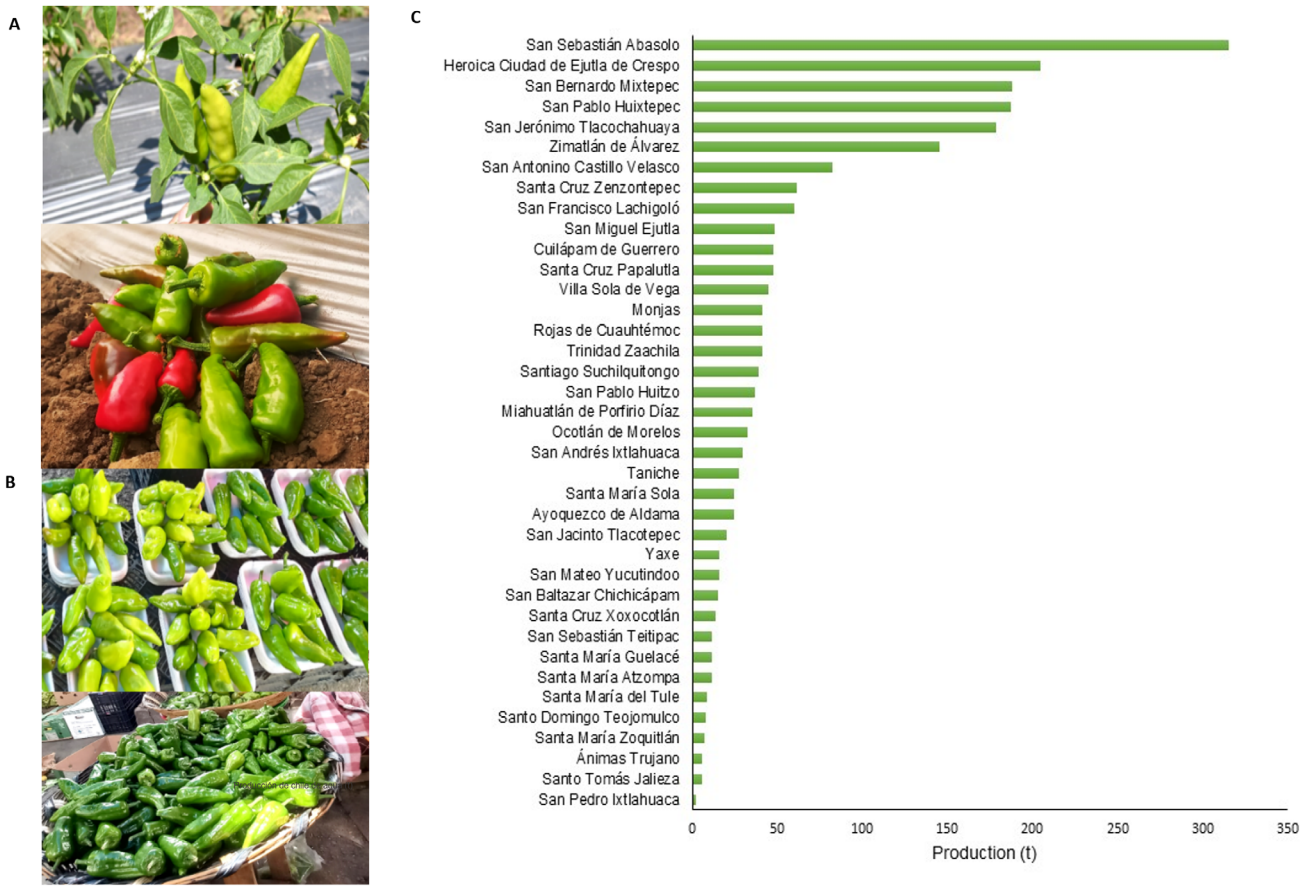


Figure 1. Chile de Agua is a type of chilli native to the Central Valleys of Oaxaca, Mexico. A, Plant and fruits at physiological maturity; B, Fruits marketed at the Oaxaca Central Market, at horticultural maturity; C, Agricultural production (t) 2023 of chile de Agua in different municipalities in the state of Oaxaca (SIAP, 2023).

1,700m elevation. The production cycle of this crop requires approximately 150–228 workdays per hectare (López-López and García-Castro, 2006; López-López and Rodríguez-Hernández, 2019), with one or two planting cycles. In 2023, 272 hectares were cultivated, with an average yield of 7.8t/ha (SIAP, 2023). However, a survey by Aparicio-del-Moral *et al*. (2013) reported yields ranging from 3.2t/ha to 5.8t/ha under open-field and gravity irrigation conditions. López-López and Rodríguez-Hernández (2019) indicate that as new technological components are introduced into the production system Feel free to change the proposed short running title as long as it stays in the 45-character (including spaces) limit Feel free to change the proposed short running title as long as it stays in the 45-character (including spaces) limit – such as artificial protection, drip fertigation and producer-selected seeds – fruit quality improves (greater proportion of first- and second-grade fruits), and yields can reach up to 16.8t/ha within a 110–122-day cycle, indicating significant opportunities for improvement.

The fruits are generally marketed at their consumption maturity stage, when they still display the characteristic light green colour. They are consumed roasted, in sauces, or stuffed; used to serve beverages such as mezcal; and even employed in floral arrangements.

Currently, regarding new varieties, only one, 'Nasha' has been developed by the National Institute for Forestry, Agricultural and Livestock Research (INIFAP) and registered in the National Catalog of Plant Varieties with breeder title No. 2465 (SNICS, 2024).

The fruit is triangular with an erect position, light green colour, thick pericarp, pointed apex and smooth skin (López-López and García-Castro, 2006). Chile de Agua is sold locally, and the unit of measurement is known as a 'carga', which contains approximately 800 top-quality chillies and weighs about 60kg. Although there is no official quality standard, quality is based on consumer demands and classified into three grades according to size, colour, brightness and absence of damage:

- First grade: fruits ≥ 10cm in length and ≥ 5cm in diameter at the base, uniform green colour, no deformities, smooth and shiny pericarp, and free from insect damage, physiological disorders or pathogens
- Second grade: fruits ≤ 10cm long and ≤ 5cm in diameter, with slight reddish discolouration, smooth and shiny pericarp, may show up to 5% damage, and can have deformities.
- Third grade: all other fruits not meeting the above criteria (Virgen, 2006; Ambrosio, 2007).

Chile Costeño

It is native to Oaxaca and is mostly cultivated in the Coast region. In 2023, agricultural production in Oaxaca was 2,415t, with yields ranging from 1.8t/ha to 3.45t/ha (Table 1; SIAP, 2023). It is distributed between the coordinates 15° 54′ 00″ N and 98° 06′ 00″ W, at elevations below 500m. The fruits are triangular without a neck at the base, with a thin, slightly wrinkled pericarp and smooth skin. This chile comes in yellow and red at physiological maturity and is sold dried, with a smooth, dark red colour (López-López and García-Castro, 2006) (Figure 2). It is used in making mole, salsas and other regional dishes (Ovando, 2007).

Table 1. 2023 agricultural production of chile Costeño in the state of Oaxaca, which describes the planted and harvested area, production and yield by municipality (SIAP, 2023).

Municipality	Area planted (ha)	Area harvested (ha)	Production (t)	Yield (t/ha)
Villa de Tututepec de Melchor Ocampo	135.20	135.20	341.65	2.53
Santiago Pinotepa Nacional	107.00	107.00	315.55	2.95
Tataltepec de Valdés	125.50	125.50	297.25	2.37
Santiago Jamiltepec	97.50	97.50	263.50	2.70
Santa María Huazolotitlán	86.50	86.50	238.80	2.76
Santo Domingo Tonalá	85.60	75.60	207.90	2.75
San Andrés Huaxpaltepec	66.45	66.45	200.25	3.01
Santa María Zacatepec	47.80	47.80	143.40	3.00
San Pedro Amuzgos	23.70	23.70	68.75	2.90
Pinotepa de Don Luis	19.50	19.50	67.30	3.45
Mártires de Tacubaya	33.25	33.25	59.85	1.80
Santo Domingo Armenta	27.00	27.00	56.70	2.10
San Antonio Tepetlapa	22.00	22.00	39.60	1.80
San Sebastián Ixcapa	17.50	17.50	37.65	2.15
San Lorenzo	15.75	15.75	34.65	2.20
Santa María Ipalapa	9.50	9.50	29.45	3.10
San Pedro Mixtepec	4.00	4.00	12.80	3.20
Total	923.75	913.75	2,415.05	2.64

Figure 2. Dried fruits of chile Costeño, native of the Coast region of Oaxaca. A, Red chile Costeño; B, Yellow chile Costeño.

Chile Soledad

It is grown in the Papaloapan region at elevations between 100 and 400m (Aguilar-Rincón *et al*, 2010). It has an average production of 1,764t and a yield of 6t/ha (Table 2; SIAP, 2023). The plant grows erect with angular stems, green with purple longitudinal stripes, and dense pubescence. The calyx lacks pigmentation. The fruits are elongated, green when immature and red when mature, with smooth skin. It is cultivated in seasonal conditions (August to March) (Aguilar-Rincón *et al*, 2010).

Table 2. 2023 agricultural production of chile Soledad in the state of Oaxaca, which describes the planted and harvested area, production and yield by municipality of Papaloapan and Itsmo region (SIAP, 2023).

Region	Municipality	Area planted (ha)	Area harvested (ha)	Prodution (t)	Yield (t/ha)
Papaloapan	San Juan Bautista Tuxtepec	94.70	81.70	530.55	6.49
Papaloapan	Loma Bonita	53.80	43.80	338.79	7.73
Papaloapan	San José Chiltepec	26.60	24.60	183.84	7.47
Papaloapan	San Lucas Ojitlán	30.00	30.00	130.5	4.35
Papaloapan	Santa María Jacatepec	17.55	15.55	109.00	7.01
Papaloapan	San Juan Cotzocón	21.50	17.00	101.7	5.98
Papaloapan	Santiago Yaveo	23.50	17.50	101.15	5.78
Papaloapan	Santiago Jocotepec	19.30	19.30	80.95	4.19
Papaloapan	San Juan Lalana	11.20	11.20	52.66	4.70
Papaloapan	Ayotzintepec	9.30	9.30	45.47	4.89
Papaloapan	San Juan Bautista Valle Nacional	6.00	6.00	42.20	7.03
Istmo	Asunción Ixtaltepec	5.25	5.25	19.69	3.75
Papaloapan	San Felipe Jalapa de Díaz	7.00	7.00	19.60	2.80
Papaloapan	San Felipe Usila	2.00	2.00	5.50	2.75
Papaloapan	San Pedro Ixcatlán	1.20	1.20	3.18	2.65
Total		328.90	291.40	1,764.78	6.06

Chile Tabaquero

Also known as chile Chiltepe, this variety is native to Oaxaca (Aguilar-Rincón *et al*, 2010; Toxqui-Tapia *et al*, 2022). It is currently cultivated in the Papaloapan región, with an average yield of 3.1t/ha (Figure 3). The plant is of intermediate growth, reaching 64.8 to 117cm in height, 120 days after transplanting. It yields an average of 97.8g per plant, with about 30 fruits averaging 3.3g and 3.6 to 4.79cm in length (Castellón-Martínez *et al*, 2014). The fruit contains higher levels of phenolic compounds (total phenols and flavonoids) when immature. It also has a low capsaicin content (6.7μg/ml) compared to other types such as piquín (116.2μg/ml), but is rich in vitamin C (Vera-Guzmán *et al*, 2011). It is consumed both fresh and dried and is used to make various dishes and sauces, including 'chintextle' (a spreadable dry chilli paste) (Aguilar-Rincón *et al*, 2010; Castellón-Martínez *et al*, 2014).

Chile Taviche

It is native to the Central Valleys of Oaxaca (Castellón-Martínez *et al*, 2014) and was one of the main ingredients and chillis in local dishes in Santo Domingo Tomaltepec up until around the 1950s. It is currently grown in Miahuatlán de Porfirio Díaz in the Central Valleys and the Sierra Sur region (Aguilar-Rincón *et al*, 2010; Sanjuan-Martínez *et al*, 2022; Sclavo-Castillo *et al*, 2024). The fruit is a triangular berry 3 to 6cm in length, mostly consumed dried to make mole and sauces known as 'chirmoles' (Figure 4) (Aguilar-Rincón *et al*, 2010). This chile is locally known but not commonly found in major Oaxaca markets. Surveys in the Benito Juárez Market, Central de Abastos in Oaxaca City, and on market days in Villa de Etla showed that vendors often confuse chile Taviche with chile Costeño, but they differ morphologically: Taviche has a distinctly triangular base. Like chile de Agua, Tusta, and Solterito, chile Taviche is favoured by consumers in the Central Valleys (Castellón-Martínez *et al*, 2012). The fresh fruits of Taviche are used to prepare sauce and green mole, while the dried fruits are used to make sauces for traditional dishes such as mole, coloradito and amarillo. Studies have been conducted on its plant and fruit characteristics based on samples collected from Santa Cruz Xitla, San Simón Almolongas, Miahuatlán de Porfirio Díaz, Ejutla de Crespo and San Miguel Ejutla (Castellón-Martínez *et al*, 2014; Sanjuan-Martínez *et al*, 2022).

Chile Huacle

It is native to the Sierra de Flores Magón region, located in the northwest of the state of Oaxaca. It is exclusively cultivated in this area, particularly in the municipalities of San Juan Bautista Cuicatlán and, more recently, in Santo Domingo del Chilar. The cultivated areas range from 0.5 to 1 hectare, with yields of 1t/ha of dried chilli (López-López and Pérez-Bennetts, 2015). The fruit is a trapezoidal-shaped berry with a smooth surface texture, a pointed apex, green when immature and dark brown at physiological maturity (Figure 5), in addition to yellow and red colour variants when ripe (López-López and Pérez-Bennetts, 2015; Galeote-Cid *et al*, 2022). It is one of the most representative chillies of Oaxaca as it is a main ingredient in the traditional 'Mole negro Oaxaqueño', internationally recognized and featured in the book *Ark of taste of Slow Food* (Slow Food, 2025); Mole negro is an emblematic dish of Oaxacan cuisine, which was

Figure 3. Chile Tabaquero cultivated in Papaloapan region of Oaxaca. A, Dried fruits of chile Tabaquero; B, 2023 agricultural production (t) of chile Tabaquero in four municipalities in Oaxaca (SIAP, 2023).

Figure 4. Dried fruits of chile Taviche, native to the Central Valleys of Oaxaca.

Figure 5. Dried fruits of chile Huacle negro, native to the Sierra de Flores Magón region of Oaxaca.

declared Intangible Cultural Heritage of the State of Oaxaca by the local Congress in 2008 and is part of Mexican cuisine, recognized as one of the finest in the world by UNESCO in 2010 (UNESCO, 2010; Baez-Vera *et al*, 2018). This chilli is highly valued and experiences peak demand during two traditional festive seasons – Day of the Dead in November, and December holidays – when its price ranges from 800 to 1,000 Mexican pesos per kilogram. Despite its cultural importance, its use is increasingly being replaced by more commercially available and affordable chillies such as ancho and guajillo.

Lesser-used cultivated chillies

This group includes domesticated chillies with limited usage, yet they play an important role in the biodiversity of Oaxacan chillies. They are cultivated on a very local scale. Notable among them is the Pasilla Oaxaca or Pasilla Mixe, native to the Mixe region of Oaxaca and cultivated in the Central Valleys, Sierra de Flores Magón, and Sierra de Juárez (Mixe region). The plant grows to about 121cm in height, with 20 fruits per plant and a yield of 263g per plant (Sanjuan-Martínez *et al*, 2022). It is mainly sold dried, with fruits smoked using oak wood in traditional ovens. It is used in mole and in making a paste known as ´chintextle´ (Aguilar-Rincón *et al*, 2010; Sanjuan-Martínez, 2020).

Tusta is cultivated in the Central Valleys, Coast, and Sierra Sur; it is collected from backyard gardens and 'milpas', and also cultivated on hillsides (Aguilar-Rincón *et al*, 2010; Taitano *et al*, 2019). At 120 days after transplanting (DAT), plants measure between 46.8cm and 60.7cm in height, with an average yield of 35.2g per plant and fruit length ranging from 1.47 to 2.8cm (Aguilar-Rincón *et al*, 2010; Castellón-Martínez *et al*, 2014; Santiago-Luna *et al*, 2016). The fruits have a short postharvest life, losing about 20% of their weight in the first three days of storage. Average capsaicin and dihydrocapsaicin contents are 51.4μg/ml and 33.5μg/ml, respectively (Vera-Guzmán *et al*, 2011; Castellón-Martínez *et al*, 2014).

In the Sierra de Flores Magón region, in addition to Huacle chilli, local chillies such as Coxle and Achilito, endemic to the area, are grown (López-López *et al*, 2016), often intercropped with Huacle (Aguilar-Rincón *et al*, 2010). Loco chilli is cultivated in the Mixteca region, though it is more commonly found in the Sierra Nevada of Puebla (Toledo-Aguilar *et al*, 2016). Onza is cultivated in the Sierra de Juárez of Oaxaca (Aguilar-Rincón *et al*, 2010). Mirador chilli is mostly grown in Veracruz and Hidalgo, while Chocolate chilli is cultivated in the state of Chiapas (Aguilar-Rincón *et al*, 2010; Ramírez *et al*, 2010).

Semiwild and wild chillies

Although there are no official records of their conservation area, wild and semiwild chillies represent an important part of the cultural and culinary identity of various Oaxacan regions. Nanche chilli (semiwild) is native to Oaxaca, found in the Central Valleys. At 120 DAT, plants reach an average height of 145.4cm, with a yield of 11.34g per plant. The fruits are about 1.35cm in length and have a short shelf life, losing 30% of fresh weight within five days (Castellón-Martínez *et al*, 2014). Solterito chilli is collected in Central Valley municipalities and shows high variability in growth traits. It grows between 94.6 and 138.8cm tall at 120 DAT, with fruit lengths ranging from 2.88 to 5.7cm (Castellón-Martínez *et al*, 2014). Other lesser-known types include Paradito or Escuchito, Soltero, de Monte, and Mirasol (with reports of Mirasol collections on the Oaxacan coast) (Pérez-Martínez *et al*, 2022).

Bolita and Piquín chillies are wild types. The former produces round fruits averaging 1.4cm in length and 1.1cm in diameter, with smooth skin. Piquín refers to various morphotypes also known as Chiltepín, Amashito, Chigolito, among others. At 120 DAT, plants reach 125 to 152.4cm in height, with fruit lengths between 1.07 and 2.26cm and diameters between 0.75 and 0.89cm (Castellón-Martínez *et al*, 2014). Both types are highly pungent (Aguilar-Rincón *et al*, 2010). Guiña Shirundu and Guiña Shuladi are also considered wild types, growing in stubble fields, milpas, backyards, pastures,

and abandoned lands in the Isthmus of Tehuantepec, Oaxaca. Guiña Shirundu fruits are round, 0.6 to 1.2cm long, with smooth skin and high pungency. Guiña Shuladi fruits are triangular, 1.8 to 3.0cm long, with smooth skin and moderate pungency. Both are known for their unique flavours in Isthmus cuisine (Aguilar-Rincón *et al*, 2010; Toledo-Martínez, 2018).

Conservation of chilli diversity in Oaxaca

Currently, chilli diversity is mainly conserved *in situ*, in the same locations where they have been domesticated, produced and cultivated by backyard or family farmers. As a result, many types are only known locally due to the small-scale cultivation for household use. However, some varieties are at risk of disappearing, including Tusta, Piquín, and Paradito, among others (Castellón-Martínez *et al*, 2012; Toxqui-Tapia *et al*, 2022).

Boege (2008) emphasizes that *in situ* conservation of the genetic diversity of *Capsicum annuum* in its centre of origin, domestication and diversification facilitates the study of genetic variation in small geographic spaces, due to strong cultural relationships. Recently, strategies such as Community Seed Banks (CSBs) have been implemented for germplasm conservation, undoubtedly including *Capsicum* spp., although the exact number of safeguarded types is unknown. CSBs emerged from the need to preserve local seeds *in situ*, directly with the producers themselves. Most were initially implemented to safeguard maize, bean and squash seeds; however, today, some also conserve other species, including chilli peppers, as part of the milpa system. The establishment of CSBs in Mexico was first coordinated by the National Service for the Inspection and Certification of Seeds (SNICS) in 2005, through institutions such as the National Institute for Forestry, Agricultural and Livestock Research (INIFAP), Chapingo Autonomous University (UACH), National Polytechnic Institute (IPN) and National Autonomous University of Mexico (UNAM), among others. In the state of Oaxaca, researchers from the INIFAP were responsible for establishing 11 CSBs in the communities of San Agustín Amatengo, San Jerónimo Coatlán, Santa Catarina Juquila, San Miguel del Puerto, San Pedro Comitancillo, Santa María Jaltianguis, Santiago Yaitepec, Santa María Peñoles, San Andrés Cabecera Nueva, and Putla de Villa Guerrero. The main activities carried out by the CSBs include: the *in situ* conservation of local crop diversity, seed selection in the field during each cropping cycle, promoting seed exchange among farmers, participation in seed fairs, information transfer through demonstration events and training sessions, and participatory breeding, among others (Aragón-Cuevas, 2016; Vera-Sánchez *et al*, 2016; SNICS, 2017). On the other hand, the National Technical Assistance Strategy Programme enables trained agricultural technicians to provide guidance and training through farmer field schools to small-scale and backyard producers, aiming to integrate traditional knowledge with scientific and technical expertise to improve productivity and support the conservation of local crops (SADER, 2020).

Over time, small-scale and backyard farmers have made significant efforts to conserve their local chilli crop. This work goes beyond simple preservation – it represents the maintenance of biological and genetic diversity in crops. Such diversity is a key resource for genetic improvement and offers an opportunity to engage primary stakeholders in participatory plant breeding. This approach considers traditional practices and customs related to cultivation, as well as farmers' preferences for preserving specific varieties based on yield, climate and soil adaptation, and cultural practices. Genetic improvement is important for all types of producers, including smallholders and backyard farmers, as they can benefit from technologies developed through breeding programmes (Castellón-Martínez *et al*, 2012; Montaño-Lugo *et al*, 2014). To strengthen the connection with farmers, a participatory plant breeding programme in chilli peppers has been launched in collaboration with vegetable crop researchers at the Campo Experimental Central Valleys of Oaxaca-INIFAP, who have developed technologies whose transfer and adoption still face significant challenges (López-López and Rodríguez-Hernández, 2019; INIFAP, 2023).

Crop breeding is not an easy task – it involves complex challenges, especially given the number of stakeholders involved, including farmers, researchers in the field, government institutions, public policies, and social organizations. Therefore, to ensure that the process is fair and beneficial to all, collaboration among all stakeholders is essential, with traditional farmer knowledge must be valued. For this reason, strong links between producers, researchers and institutions are crucial. Working together, they can contribute to breeding programmes, the creation or strengthening of community seedbanks, seed exchange networks, and other initiatives that promote the conservation of crop diversity, improve production and respect the ancestral knowledge and traditions of local communities.

Most local Oaxaca chilli varieties are preserved *ex situ* in orthodox seed germplasm banks. These include collections at the Orthodox Seeds from the South collection located at the Campo Experimental Centrall Valleys of Oaxaca (part of Regional Research Center, Pacific South of INIFAP), at the Autonomous University of Chapingo – Southern Regional University Center (UACH-CRUS) in Oaxaca, and accessions at the National Center for Genetic Resources (CNRG) in Jalisco, collected mainly from Chiapas, Guerrero, Veracruz and especially Oaxaca. However, further research programmes are needed to continue exploring, conserving and improving this resource.

In the National Catalogue of Plant Varieties, of the 97 crops listed, only 2% are chilli varieties, most of them commercial germplasm such as Ancho, Guajillo, Jalapeño, Puya, Pimiento, Habanero, and Manzano; only one improved variety of native Oaxacan chilli, chile de Agua, called 'Nasha,' is registered (SNICS, 2024). This indicates a clear opportunity to implement genetic improvement programmes as a strategy for use and conservation.

Although all chilli varieties in Oaxaca possess significant agronomic and cultural value, it is evident that some have gained greater visibility and persistence, while others are at risk of disappearing. This situation cannot be explained solely by their agronomic or culinary characteristics, but also by external factors that affect their cultivation and intergenerational transmission. Among these factors are the economic volatility of rural areas, which forces many families to migrate or shift to more profitable crops; agricultural policies that have historically favoured large-scale production and commercial varieties; and the progressive homogenization of diets, influenced by market forces and urban food habits, which reduces the demand for lesser-known local varieties. Additionally, climate change poses an increasing threat, as it

alters agricultural cycles and water availability, particularly affecting traditional varieties adapted to specific conditions (Pérez-Martínez *et al*, 2022; Toxqui-Tapia *et al*, 2022; Soleri *et al*, 2023). Therefore, the conservation of chilli diversity in Oaxaca requires not only agronomic and genetic actions, but also a comprehensive approach that takes into account economic, sociocultural and environmental contexts.

Challenges and opportunities for chilli cultivation

Genetic diversity in chilli peppers is protected by maintaining the production of local varieties that possess unique characteristics – such as colour, flavour, resistances or specific culinary uses. Most chilli farmers in Oaxaca currently use traditional open-field methods with agrochemicals and gravity-fed irrigation (Aparicio-del-Moral *et al*, 2013; López-López *et al*, 2016). Also small-scale chilli producers in Oaxaca commonly use agrochemicals and gravity-fed irrigation; however, it is important to note that backyard producers generally rely on the rainy season. Pérez-Acevedo *et al*, (2017) surveyed chile de Agua producers in the Central Valleys during 2013 and 2014, reporting that the majority of the 63 farmers interviewed cultivated half a hectare or less. Regionally, chile de Agua is the most widely grown local variety. Fertilization is carried out either organically or through a combination of organic and inorganic inputs, and pesticides are commonly used. Some of the challenges they face, which sometimes lead to the abandonment of the crop, are:

- Incidence of pests and disease. Diseases include: wilt of chillies (associated with *Rhizoctonia, Fusarium*, *Phytophthora* and *Pythium*), virus diseases and foliage diseases *(Altenaria* spp. and *Cercospora* spp., among others). Pests: Fruit weevil, chilli beetle, insect vectors of viruses (López-López and Castro-García 2006; Pérez-Acevedo *et al*, 2017).
- Untapped productive potential. In chile de Agua, yields range from 3.2t/ha to 7.6t/ha, including fruits considered first, second and third quality due to their size, and in chilli Huacle, yields of 1t/ha in dry weight (Aparicio-del-Moral *et al*. 2013; López-López and Pérez-Bennetts, 2015; SIAP, 2023).
- High production costs. Mainly due to the use of agrochemicals and fertilizers (Aparicio-del-Moral *et al*, 2013).
- Inefficient water use, due to the irrigation system used (gravity) (Aparicio-del-Moral *et al*. 2013; López-López *et al*, 2016).
- Postharvest losses, including drying of fruit in direct sunlight (increased pathogen damage, fruit staining and loss of compounds) and short shelf life of fresh fruit (moisture loss of at least 15% after 4 d of storage in chilli Tusta, Tabaquero, Piquín and Nanche) (Aparicio-del-Moral *et al*, 2013; Castellón-Martínez *et al*, 2014; López-López and Pérez-Bennetts, 2015).

According to previous work carried out on chilli cultivation, potential research opportunities include:

- Rescue and maintain the diversity of chilli morphotypes through breeding programmes such as participatory breeding that involve key stakeholders (farmers) and allow *in situ* conservation (Sánchez-Hernández *et al*, 2016; Salgotra and Chauhan, 2023).
- Implement technologies to make water use more efficient, by using a more appropriate irrigation system, such as drip irrigation, which reduces water consumption by up to 50% (SADER, 2024), and using mulch on the soil, which limits evapotranspiration (El-Beltagi *et al*, 2022).
- Apply organic sources. These improve agronomic characteristics such as yield, plant height, flowering time, among other characteristics such as the nutraceutical quality of the fruits (Márquez-Hernández *et al*, 2013). For example, the use of compost with application of *Azospirillum* sp. rhizobacteria in chilli Huacle grown under greenhouse, had similar yields, without statistical differences compared to the control (conventional management) (4,122kg/ha and 4,464kg/ha, respectively) (Galeote-Cid *et al*, 2022). In addition, the use of organic sources reduces production costs.
- Use production technologies and postharvest management: (1) different planting densities; (2) growing the germplasm under protected conditions – for example, in chile de Agua, if macro-tunnels are used, the yield increases 300-600% more than growing it in open air (Escamirosa-Tinoco *et al*, 2021); (3) drying in electric or solar dryers, which reduce drying time and prevent contamination by avoiding exposure to open air; in addition, the type of drying allows maintaining or degrading bioactive compounds such as vitamin C and phenolic and carotenoid compounds (Montoya-Ballesteros *et al*, 2014; Castillo-Téllez *et al*, 2018); (4) pretreatments in drying, such as blanching (95 °C for 3 min), which can be applied to chillies – as has been done in chile de Agua – to reduce drying time and oxidation of organic compounds (Bautista *et al*, 2021). If producers are able to reduce post-harvest losses, they will have greater incentives to continue cultivating traditional varieties, which may have lower yields but hold high cultural and gastronomic value.

This is why collaborations among scientists, local farmers, cooks and merchants are essential for the conservation of Oaxaca's endemic chilli varieties, as each of these actors contributes distinct and complementary knowledge, experiences, and strategies. Involving cooks and merchants helps ensure that chilli peppers continue to be used and sold, which creates demand and prevents their abandonment. For example, if a chilli variety such as chile Huacle or chile de Agua continues to have a valued and profitable culinary use, its cultivation will likely be maintained. The collective work of all stakeholders can contribute to revaluing these chilli varieties as part of Oaxaca's cultural identity and promote their local consumption. On the other hand, scientists can collaborate with farmers to adapt agricultural practices that protect local varieties from climate change, pests or genetic erosion. Moreover, such collaborations enable the development of educational projects or awareness campaigns highlighting the importance of native chilli varieties.

Conclusions

Mexico is home to a great diversity of chilli peppers,

which are a fundamental part of the daily diet of its population. The chilli genetic resources available in Oaxaca are known at the local level but remain largely unknown to the broader population across the various regions of the state, especially outside the areas where they are cultivated. This is primarily because most of these varieties are native, semi-domesticated or wild. Oaxaca is the state with the greatest diversity, represented by at least 25 types distributed across agroecological niches where they are still preserved. Therefore, it is essential to continue efforts in their conservation and sustainable use.

The importance of conserving genetic resources lies in ensuring access to genetic material that can be used in research, enabling deeper understanding and scientific advancement. At the same time, it can be incorporated into breeding programmes to help develop alternatives to address current challenges faced by chilli cultivation – whether biotic or abiotic factors – many of which stem from vulnerabilities caused by climate change. This enables agriculture to adapt to changing conditions and contributes to food security. The conservation of chilli pepper diversity in Oaxaca largely depends on the work of small-scale producers, who, through their traditional agricultural practices and local knowledge, have managed to maintain a wide variety of chilli types over time. This conservation is not solely genetic or agronomic; it is deeply interwoven with the culinary and cultural traditions of the state's diverse regions. In many Oaxacan communities, the cultivation, use and selection of specific chilli varieties are tied to food practices passed down through generations, where each type of chilli is used according to its flavour, pungency, colour, texture or symbolic value. Therefore, the conservation of chilli diversity in Oaxaca cannot be understood without recognizing the central role of local producers as custodians of biocultural knowledge.

Acknowledgments

We also acknowledge the support provided by the researchers Luis Eduardo García Mayoral, Miguel Ángel Cano García, Filemón Rafael Rodríguez Hernández, Salvador Montes Hernández and Leodegario Osorio Alcalá, whose help allowed us to complete the description of the environment in this work.

Author contributions

Yeimy Clemencia Ramírez-Rodas prepared the study proposal, collected and organized the data and information, and wrote the manuscript. Luis Yobani Gayosso-Rosales contributed to the review and improvement of the study proposal, data collection, writing, editing and improving drafts of the manuscript. Ulises Santiago-López contributed to the review and improvement of the study proposal, editing and improving drafts of the manuscript.

Conflict of interest statement

The authors declare no conflicts of interest.

References

Aguilar-Meléndez, A., Lira Noriega, A. (2018). ¿Dónde crecen los chiles en México?. En: *Los chiles que le dan sabor al mundo*. Eds. A. Aguilar-Meléndez, M. A. Vásquez-Dávila, E. Katz y M. R. Hernández Colorado editores (Universidad Veracruzana, México y IRD, Francia) 75–92.

Aguilar-Rincón V H, T Corona-Torres, P López-López, L LatournerieMoreno, M. Ramírez-Meraz, H Villalón-Mendoza, J A AguilarCastillo. (2010). Los Chiles de México y su distribución (SINAREFI, Colegio de Postgraduados, INIFAP, IT-Conkal, UNAL y UAN. Estado de México, México), 108 p.

Ambrosio S., F. (2007). Análisis de los aspectos socio-culturales y económicos del agroecosistema chile de agua (*Capsicum annuum* L.) en Cuilapam de Guerrero, Oaxaca. Tesis de Maestría, Instituto Tecnológico del Valle de Oaxaca, México.

Aparicio-del-Moral, J. O., Tornero-Campante, M. A., Sandoval-Castro, E., Villarreal-Manzo, L. A., de los Ángeles Rodríguez-Mendoza, M. (2013). Factores sociales y económicos del cultivo de chile de agua (*Capsicum annum* L.) en tres municipios de los valles centrales de Oaxaca. *Ra Ximhai* 9(1), 17–24.

Aragón C., F. (2016). México: Bancos comunitarios de semillas en Oaxaca. En: Bancos Comunitarios de Semillas: Orígenes, Evolución y Perspectivas, eds: Vernooy R., P Shrestha., B Sthapit,. Y M Ramírez., (Bioversity International, Lima Perú), 136–139.

Báez-Vera, P. J., Malda Barrera, G., Zocchi, D. M. (2018). El Arca del Gusto en México: Productos, saberes e historias del patrimonio gastronómico (Slow Food Editore, Bra., Italy), 181p .

Bautista, P. B., Juárez, N. A., Hernández, I. H. (2021). Efecto del escalde en el secado de chile de agua producido en la región de los valles centrales de Oaxaca. *Universidad & ciencia* 10, 50–64.

Boege, S.E. (2008). El patrimonio biocultural de los pueblos indígenas de México (Instituto Nacional de Antropología e Historia y Comisión Nacional para el Desarrollo de los Pueblos Indígenas, Ciudad de México, México), 344 p.

Castellón-Martínez E., Chávez-Servia, J. L., Carrillo-Rodríguez, J. C., Vera-Guzman, A. M. (2012). Preferencias de consumo de chiles (*Capsicum annuum* L.) nativos en los Valles Centrales de Oaxaca, México. *Revista Fitotecnia Mexicana* 35 (5), 27–35.

Castellón-Martínez, E., Carrillo-Rodríguez, J. C., Chavez-Servia, J. L., Vera-Guzmán, A. M. (2014). Variación fenotípica de morfotipos de chile (*Capsicum annuum* L.) nativo de Oaxaca, México. *Phyton* 83(2), 225–236.

Castillo-Téllez, M. C., Figueroa, I. P., Téllez, B. C., Vidaña, E. C. L., Ortiz, A. L. (2018). Solar drying of Stevia (*Rebaudiana Bertoni*) leaves using direct and indirect technologies. *Solar Energy*, *159*, 898–907. doi: https://doi.org/10.1016/j.solener.2017.11.031

El-Beltagi, H. S., Basit, A., Mohamed, H. I., Ali, I., Ullah, S., Kamel, E. A., Shalaby, T. A., Ramada, K. M. A., Alkhateeb A. A., Ghazzawy, H. S. (2022). Mulching as a sustainable water and soil saving practice in agriculture: A review. *Agronomy* 12(8):1881. doi: https://doi.org/10.3390/agronomy12081881.

Escamirosa-Tinoco, C., Martínez-Gutiérrez, G. A., Morales, I., Aquino-Bolaños, T., Cortés-Martínez, C. I., Cruz-Andrés, O. R. (2021). Rendimiento de chile de agua bajo diferentes cubiertas de macrotúnel. Revista fitotecnia mexicana 44(3), 333–340.

FAO (2012). Segundo plan de acción mundial para los recursos fitogenéticos para la alimentación y la agricultura. FAO-Comisión de Recursos Genéticos para la Alimentación y la Agricultura. 24 p. Url: https://www.fao.org/4/i2650s/

i2650s.pdf.
Galeote-Cid, G., Cano-Ríos, P., Ramírez-Ibarra, J. A., Nava-Camberos, U., Reyes-Carrillo, J. L., Cervantes-Vázquez, M. G. (2022). Comportamiento del chile Huacle (*Capsicum annuum* L.) con aplicación de compost y Azospirillum sp. en invernadero. Terra Latinoamericana 40:1–12 e828. doi: https://doi.org/10.28940/terra.v40i0.828.
Guzmán-Mendoza, R., Hernández-Hernández, V., Salas-Araiza, M. D., Núñez-Palenius, H. G. (2022). Diversidad de especies de plantas arvenses en tres monocultivos del Bajío, México. Polibotánica, (53), 69–85.
INIFAP (2023). Instituto Nacional de Investigaciones Forestales Agrícolas y Pecuarias. Tecnología para la producción de chile de agua en ambiente protegido. Url: https://www.gob.mx/inifap/articulos/tecnologia-para-la-produccion-de-chile-de-agua-en-ambiente-protegido?idiom=es
López-López, P., Castro-García, F. H. (2006). La diversidad de los chiles (*Capsicum* spp., Solanaceae) de Oaxaca. En: Avances de investigación de la red de hortalizas del SINAREFI , eds. López L. P y S. Montes H. (Instituto Nacional de Investigaciones Forestales, Agrícolas y Pecuarias, Campo Experimental Bajío. Celaya, Gto. México), 135–178.
López-López, P., Pérez-Bennetts, D. (2015). El chile huacle (*Capsicum annuum* sp.) en el estado de Oaxaca, México. *Agroproductividad* 8(1), 37–39.
López-López, P., Rodríguez-Hernández, R. (2019). Evaluación de chile de agua (*Capsicum annuum* L.) en ambiente protegido y poda en los valles centrales de Oaxaca. *Producción Agropecuaria: Un enfoque integrado* 47– 52.
López-López, P., Rodríguez-Hernández, R., Bravo-Mosqueda, E. (2016). Impacto económico del chile huacle (*Capsicum annuum* L) en el estado de Oaxaca. *Revista Mexicana de Agronegocios* 38, 317–327.
Márquez-Hernández, C., Cano-Ríos, P., Figueroa-Viramontes, U., Avila-Diaz, J. A., Rodríguez-Dimas, N., García-Hernández, J. L. (2013). Rendimiento y calidad de tomate con fuentes orgánicas de fertilización en invernadero. *Phyton* 82(1), 55–61.
Martínez-Martínez, R., Méndez-Infante, I., Castañeda-Aldaz, H. M., Vera-Guzmán, A. M., Chávez-Servia, J. L., Carrillo-Rodríguez, J. C. (2014). Heterosis interpoblacional en agromorfología y capsaicinoides de chiles nativos de Oaxaca. *Revista Fitotecnia Mexicana* 37(3), 199–207.
Montaño-Lugo, M. L., Velasco Velasco, V. A., Ruíz Luna, J., Campos Ángeles, G. V., Rodríguez Ortiz, G., Martínez Martínez, L. (2014). Contribución al conocimiento etnobotánico del chile de agua (*Capsicum annuum* L.) en los Valles Centrales de Oaxaca, México. Revista mexicana de ciencias agrícolas, 5(3), 503–511.
Montoya-Ballesteros, L. C., González-León, A., García-Alvarado, M. A., Rodríguez-Jimenes, G. C. (2014). Bioactive compounds during drying of chili peppers. *Drying Technology* 32(12), 1486–1499. doi: https://doi.org/10.1080/07373937.2014.902381.
Ovando, CME. (2007). Producción de chile costeño con agribón y fertirriego en la costa de Oaxaca. *Revista Agroproduce,* 20–21.
Pérez-Acevedo, C. E., Carrillo-Rodríguez, J. C., Chávez-Servia, J. L., Perales-Segovia, C., Enríquez del Valle, R., Villegas-Aparicio, Y. (2017). Diagnóstico de síntomas y patógenos asociados con marchitez del chile en Valles Centrales de Oaxaca. *Revista mexicana de ciencias agrícolas* 8(2), 281–293. doi: https://doi.org/10.29312/remexca.v8i2.50.
Pérez-Martínez, A. L., Eguiarte, L. E., Mercer, K. L., Martínez-Ainsworth, N. E., McHale, L., van der Knaap, E., Jardón-Barbolla, L. (2022). Genetic diversity, gene flow, and differentiation among wild, semiwild, and landrace chile pepper (*Capsicum annuum*) populations in Oaxaca, Mexico. *American Journal of Botany* 109(7), 1157–1176. doi: https://doi.org/10.1002/ajb2.16019.
Pickersgill, B. 2016. Chile Peppers (*Capsicum* spp.). En: Ethnobotany of Mexico. Interactions, eds. Lira, R., Casas, A., Blancas, J., (Springer Science + Business Media, New York, NY), 417–437.
Ramírez, H., Amado-Ramírez, C., Benavides-Mendoza, A., Robledo-Torres, V., Martínez-Osorio, A. (2010). Prohexadiona-Ca, AG3, ANOXA y BA modifican indicadores fisiológicos y bioquímicos en chile Mirador. *Revista Chapingo. Serie horticultura* 16(2), 83–89.
Rodríguez-Campos, E. (2018). La diversidad genética de *Capsicum annum* de México. En: Los chiles que le dan sabor al mundo, eds. A. Aguilar-Meléndez, M. A. Vásquez-Dávila, E. Katz y M. R. Hernández Colorado editores, (Universidad Veracruzana, México y IRD Éditions, Francia), 52–67.
Ruiz-Núñez, N. d. C.; Vásquez-Dávila, M.A. Etnoecología del chile de campo en Guelavía, Oaxaca. 2018. En: Los chiles que le dan sabor al mund, eds. A. Aguilar-Meléndez, M. A. Vásquez-Dávila, E. Katz y M. R. Hernández Colorado editores, (Universidad Veracruzana, México y IRD Éditions, Francia), 260–280.
SADER (2020). Secretaría de Agricultura y Desarrollo Rural. Estrategia de Acompañamiento Técnico. Url: https://www.gob.mx/agricultura/articulos/estrategia-de-acompanamiento-tecnico.
SADER (2024). Secretaría de Agricultura y Desarrollo Rural. Maximizando la eficiencia agrícola: Sistema de riego por goteo. Url: https://www.gob.mx/agricultura/articulos/maximizando-la-eficiencia-agricola-sistema-de-riego-por-goteo
Salgotra, R. K., Chauhan, B. S. (2023). Genetic diversity, conservation, and utilization of plant genetic resources. *Genes*, 14(1), 174. doi: https://doi.org/10.3390/genes14010174.
Sánchez-Hernández, C., Sánchez-Hernández, M. A., González-Montiel, L., Vicente-Pinacho, A. J. (2016). Mejoramiento participativo del chile huacle en Cuicatlán, Oaxaca, México. En: Conocimiento multidisciplinario. Hablando de emprendurismo educación y derecho, ed. B. López Azamar, (Universidad del Papaloapan. Oaxaca, México), 135–149.
Sanjuan Martínez, J. Características vegetativas y de frutos de tres tipos de chiles de oaxaca producidos en invernadero. (2020). *Contribución al Conocimiento Científico y Tecnológico en Oaxaca* 4(4), 50–59.
SanJuan Martinez, J., Ortiz Hernández, Y. D., Aquino Bolaños, T., Cruz Izquierdo, S., Pérez Pacheco, R. Evaluación *ex situ* de seis chiles nativos de Oaxaca en invernadero. (2022). *Revista de Ciencias Agronómicas Aplicadas y Biotecnología* 2, 101–107.
Santiago-Luna, E. G., Carrillo-Rodríguez, J. C., Chávez-Servia, J. L., del Valle, R. E., Villegas-Aparicio, Y. (2016). Phenotypic variation in tusta pepper populations. *Agronomía Mesoamericana* 27, 139–149. doi: https://doi.org/10.15517/am.v27i1.21893.
Sclavo-Castillo, D., Volkow, L. P., Martínez, E. H. (2024). El chile tabiche: memoria, reencuentro y acción comunitaria en Santo Domingo Tomaltepec, Oaxaca, México. *Naturaleza y Sociedad. Desafíos Medioambientales* 8, 79–103. doi: https://doi.org/10.53010/NLTT9430.
SIAP (2023). Anuario Estadístico de la Producción Agrícola. Url: https://www.gob.mx/siap/acciones-y-programas/produccion-agricola-33119

Slow Food (2025). Url: https://www.fondazioneslowfood.com/en/ark-of-taste-slow-food/chilhuacle/

SNICS (2017). Bancos Comunitarios de Semillas como estrategia de Conservación *in situ*. url: https://www.gob.mx/snics/acciones-y-programas/bancos-comunitarios-de-semillas-como-estrategia-de-conservacion-in-situ

SNICS (2024). Catálogo Nacional de Variedades Vegetales. Url: https://www.gob.mx/snics/es/articulos/catalogo-nacional-de-variedades-vegetales-en-linea

Soleri, D., Cleveland, D. A., Aragón Cuevas, F., Jimenez, V., Wang, M. C. (2023). Traditional foods, globalization, migration, and public and planetary health: the case of tejate, a maize and cacao beverage in Oaxacalifornia. *Challenges*, 14(1), 9.

Taitano, N., Bernau, V., Jardón-Barbolla, L., Leckie, B., Mazourek, M., Mercer, K., McHale, L., Michel, A., Baumler, D., Kantar, M., van der Knaap, E. (2019). Genome-wide genotyping of a novel Mexican chile pepper collection illuminates the history of landrace differentiation after *Capsicum annuum* L. domestication. *Evolutionary Applications* 12, 78–92. doi: https://doi.org/10.1111/eva.12651.

Toledo-Aguilar, R., López-Sánchez, H., López, P. A., Guerrero-Rodríguez, J. D. D., Santacruz-Varela, A., Huerta-de la Peña, A. (2016). Diversidad morfológica de poblaciones nativas de chile poblano. *Revista mexicana de ciencias agrícolas* 7(5), 1005–1015.

Toledo-Martínez, A. El shigundu, uno de los sabores de la cocina istmeña. (2018). En: Los chiles que le dan sabor al mundo, eds. A. Aguilar-Meléndez, M. A. Vásquez-Dávila, E. Katz y M. R. Hernández Colorado editores, (Universidad Veracruzana, México y IRD Éditions, Francia), 68–74.

Toxqui-Tapia, R., Peñaloza-Ramírez, J. M., Pacheco-Olvera, A., Albarran-Lara, L., Oyama, K. (2022). Genetic diversity and genetic structure of *Capsicum annuum* L., from wild, backyard and cultivated populations in a heterogeneous environment in Oaxaca, Mexico. *Polibotánica* 53, 87–103. doi: https://doi.org/10.18387/polibotanica.53.6.

UNAM (2007). *Capsicum rhomboideum* (Dunal) Kuntze. Herbario Nacional de México (MEXU), Plantas Vasculares. Departamento de Botánica, Instituto de Biología de la Universidad Autónoma de México. url: https://datosabiertos.unam.mx/IBUNAM:MEXU:1242472.

UNESCO (2010). La cocina tradicional mexicana: Una cultura comunitaria, ancestral y viva y el paradigma de Michoacán. Url: https://ich.unesco.org/es/RL/la-cocina-tradicional-mexicana-una-cultura-comunitaria-ancestral-y-viva-y-el-paradigma-de-michoacn-004009

Vera-Guzmán, A. M., Chávez-Servia, J. L., Carrillo-Rodríguez, J. C., López, M. G. (2011). Phytochemical evaluation of wild and cultivated pepper (*Capsicum annuum* L. and C. *pubescens* Ruiz & Pav.) from Oaxaca, Mexico. *Chilean journal of agricultural research* 71(4), 578.

Vera-Sánchez, K. S, R. González S, F. Aragón C. (2016). Bancos comunitarios de semillas en México: Una estrategia de conservación *in situ*. En: Bancos Comunitarios de Semillas: Orígenes, Evolución y Perspectivas, eds. Vernooy R., P Shrestha., B Sthapit,. Y M Ramírez, (Bioversity International, Lima Perú), 248–253.

Virgen J., S. D. (2006). Rentabilidad y mercadeo de la producción de chile de agua bajo invernadero en el municipio de Ayoquezco de Aldama, Oaxaca. Tesis de Licenciatura, Departamento de Fitotecnia.Universidad Autónoma Chapingo.Texcoco, México.

Genetic Resources

ORIGINAL ARTICLE

Genetic Resources (2026), 7 (13), 12–28
DOI: 10.46265/genresj.PUGG4952
https://www.genresj.org
ISSN: 2708-3764

Multivariate analysis of morpho-biometric diversity in indigenous chickens from two zones of Oromia Regional State, Ethiopia

Fikrineh Negash*[,a] and Usman Abdulkadir[b]

[a] *Freelance Researcher and Consultant, Addis Ababa, Ethiopia*
[b] *Adami Tulu Agricultural Research Center, P. O. Box 35, Batu, Ethiopia*

Abstract: This study characterized the morpho-biometric features of indigenous chickens across six districts of West Arsi and East Shoa zones of Oromia Region State, Ethiopia, using multivariate techniques. Data were collected from 621 mature chickens (134 males, 487 females). Univariate analysis revealed significant district-level variation in morphometric traits ($p < 0.001$), with Lume chickens exhibiting the highest values, and Siraro and Shashemene the lowest. Shank length and body weight had the highest model explanatory power (R^2: 0.58–0.64). Qualitative traits displayed distinct patterns, with red and brown plumage predominating in males and females, respectively, while rose and single combs, white–red earlobes, white skin, plain head shapes, and yellow shanks were the most prevalent. Multiple correspondence analysis highlighted associations between qualitative traits and districts, with the first two dimensions explaining 70.23% of the variance. Quadratic discriminant analysis classified chickens into their districts of origin with 41.80–91.30% accuracy, which was highest for Bora and Lume. Stepwise discriminant analysis identified seven traits (females) and four (males) as key discriminators, while canonical discriminant analysis revealed that the first two functions explained 98% of the variance in both female and male chickens, with strong between-district differentiation. Biplots confirmed that East Shoa chickens (larger in size) clustered separately from West Arsi populations. These findings underscore the phenotypic diversity of Ethiopian indigenous chickens, which is likely shaped by genetic, environmental and cultural factors. This diversity offers opportunities for targeted breeding and conservation programmes. Future studies should integrate genetic analyses to elucidate admixture and enhance breeding strategies.

Keywords: East Shoa zone, Morphometric traits, Multivariate analysis, Phenotypic diversity, West Arsi zone

Citation: Negash, F. and Abdulkadir, U. (2026) "Multivariate analysis of morpho-biometric diversity in indigenous chickens from two zones of Oromia Regional State, Ethiopia", *Genetic Resources*, 7(13), pp. 12–28. doi: 10.46265/genresj.PUGG4952.

Introduction

Poultry production serves as a critical component of Ethiopia's livestock system, playing a pivotal role in enhancing food security, generating income, and improving livelihoods for smallholder farmers. According to recent statistics (FAO, 2024), the country's poultry population stands at approximately 55 million chickens, with indigenous breeds being predominant. Despite their low productivity compared with commercial breeds, indigenous chickens are well-adapted to local environmental challenges. They thrive under low-input management systems, enduring poor feed resources, harsh climatic conditions and prevalent disease pressures. These adaptations derive from hereditary traits that produce varied responses to environmental stimuli, closely tied to anatomical-physiological features developed through natural selection (Ngeno *et al*, 2014). This evolutionary process has resulted in a wide genetic diversity within indigenous chicken populations.

The phenotypic diversity observed in Ethiopian indigenous chicken populations (Dana *et al*, 2010; Melesse and Negesse, 2011; Moreda *et al*, 2014; Negassa *et al*, 2014; Getachew *et al*, 2016; Tareke *et al*, 2018; Bekele *et al*, 2021; Mustefa *et al*, 2021; Mekonnen *et al*, 2023; Muluneh *et al*, 2023; Belay *et al*, 2024; Chebo *et al*, 2024; Markos *et al*, 2024; Begna *et al*, 2025) reflects their adaptation to various agroecological conditions. The remarkable diversity in indigenous

* Corresponding author: Fikrineh Negash (fikrineh2010@gmail.com)

Received: 03.06.2025 | Accepted: 28.11.2025 | Published online: 22.01.2026

chicken genetic resources likely stems from heterogeneous agroecological zones, diverse climatic conditions and distinct poultry-keeping practices shaped by varying production objectives (Dessie *et al*, 2011; Lawal and Hanotte, 2021). These chickens have developed both distinctive phenotypes and unique genetic profiles through generations of targeted breeding by local communities and natural selection (Mekonnen *et al*, 2023). Moreover, the observed diversity has been attributed to ethnic and cultural influences, historical migration patterns, and the country's strategic location in the Horn of Africa, serving as a crossroads between Asia and the Western world (Hassen *et al*, 2007).

This diversity serves as a vital foundation for breeding programmes aimed at improving productivity while preserving adaptive traits. It also plays a crucial role in the food security and livelihoods of marginalized farmers (Cabarles *et al*, 2012). However, these valuable genetic resources face mounting threats from shifting production systems, uncontrolled crossbreeding, environmental degradation and natural disasters (Besbes, 2009). Such pressures risk permanent genetic erosion, particularly in areas where exotic chicken breeds are being introduced. Since the loss of indigenous genetic resources is irreversible, urgent conservation measures are needed to safeguard these populations and their unique traits (Liyanage *et al*, 2015). The characterization of indigenous chickens – including their production environments and management systems – should serve as a fundamental requirement for developing sustainable conservation and utilization strategies as well as genetic improvement programmes for these genetic resources (Yussif *et al*, 2023; Liswaniso *et al*, 2024).

Both quantitative and qualitative morphological traits provide valuable tools for assessing genetic diversity in indigenous chickens (Getachew *et al*, 2016). This characterization yields critical information about current utilization potential while documenting population status and evaluating extinction risk (Tixier-Boichard *et al*, 2008; FAO, 2012). While numerous phenotypic characterization studies have been conducted across Ethiopia, a notable gap remains in multivariate analyses of chicken populations. This study specifically addresses this gap by employing multivariate statistical techniques to examine the morphological and biometric traits of indigenous chickens in selected districts of the West Arsi and East Shewa zones of the Oromia Regional State. This research aims to identify distinct phenotypic variations that will inform strategic breeding approaches and promote sustainable utilization of Ethiopia's valuable indigenous chicken genetic resources.

Materials and methods

Study area

This study was conducted in three districts, each from the East Shoa (i.e. Adama, Bora and Lume) and West Arsi (i.e. Dodola, Shashemene and Siraro) zones of Oromia Regional State, located in central and south-central Ethiopia. The districts were purposively chosen because of their socioeconomic importance in poultry production and high populations of indigenous chickens. Data were collected from randomly selected smallholder farmers who reared exclusively indigenous, non-descript chicken types – traditional breeds naturally adapted to local conditions.

Data generation

Morphometric measurements and qualitative morphological features were collected from 621 mature indigenous chickens (134 males and 487 females) aged eight months or older, an age threshold defined by Dana *et al* (2010). The number of male chickens from Adama, Lume, Bora, Dodola, Shashemene and Siraro were 25, 22, 23, 21, 21 and 22, respectively, while the corresponding numbers of female chickens were 79, 83, 79, 81, 81 and 84. Age was determined through farmer recall, and this threshold was selected because indigenous chickens are known to mature slowly (Melesse and Negesse, 2011). Sampling mature animals was also necessary due to the age-dependent expression and environmentally sensitive nature of most quantitative traits (FAO, 2012). Body weight was measured using a hanging spring balance, while linear measurements were taken with a textile measuring tape.

All the measurements followed standardized protocols (FAO, 2012):

- Body length (BL): Distance from the beak tip (*rostrum maxillae*) to the tail base (*cauda*; excluding tail feathers)
- Chest circumference (CC): Girth at the deepest point of the breast
- Shank length (SL): *Tarsometatarsus* length, measured from the flexed hock joint to the spur base
- Shank circumference (SC): Circumference at the midpoint of the shank
- Wing span (WS): Distance between the tips of both fully extended wings
- Wattles length (WL): Linear measurement from the wattle's origin to its distal tip
- Comb length (CL): Distance from the comb's anterior insertion (near the beak) to the posterior tip of the largest lobe.

Qualitative traits, including feather distribution and morphology, body plumage colour, shank colour, shank feather (absent/present), skin colour, earlobe colour, earlobe presence, comb type, head shape and spur presence, were also evaluated on the basis of standardized descriptors used in previous studies (Dana *et al*, 2010; Melesse and Negesse, 2011).

Statistical analysis

The data were analyzed using various procedures in JMP Pro 17.0.0 (SAS Institute Inc., 2022). First, separate one-way ANOVAs were conducted for male and female chickens to examine differences in morphometric variables among chickens from the six districts. Significantly different means were compared using Tukey's HSD test at $p \leq 0.05$. The statistical model used was:

$$Y_{ij} = \mu + A_i + e_{ij}$$

Where Y_{ij} is the individual morphometric or morphological measurement; μ is the overall mean; A_i is the fixed effect of the district (i = 1 to 6); and e_{ij} is the random error.

Second, contingency analysis incorporating frequency distributions (percentages) and Pearson's chi-square (χ^2) tests were performed to assess qualitative morphological variables. The χ^2 test was employed to assess the degree of

variation in these traits across the six districts. Third, multiple correspondence analysis (MCA) was used to visualize associations among categorical morphological traits and identify variation patterns.

Finally, discriminant analysis techniques – including quadratic discriminant analysis (QDA), stepwise discriminant analysis (SDA) and canonical discriminant analysis (CDA) – were applied (JMP Statistical Discovery LLC, 2022) to discriminate chickens from the six districts on the basis of morphometric variables. The original eight traits were analyzed using QDA separately for males and females, with classification accuracy determined as the percentage of individuals correctly assigned to their respective districts. The SDA was then applied to these eight traits to select the most discriminative ones. The selected traits were subsequently analyzed using CDA to evaluate morphometric differentiation among district populations. CDA drives set new variables, called canonical functions (CAN), from linear combinations of the original variables (Conte *et al*, 2018). These functions are designed to maximize the discrimination among the classes specified by the categorical grouping variable (i.e. district).

Results

Univariate analysis

Table 1 presents descriptive statistics for morphometric variables of indigenous chickens sampled across six districts, including univariate test results. All traits showed highly significant variation ($p < 0.001$) among districts for both sexes. The difference was also observed between male and female birds. Chickens from Lume district consistently exhibited the highest values for most traits, while those from Siraro and Shashemene districts showed the lowest values. Notably, Lume chickens recorded the highest mean body weights (1854.55g for males and 1435.37g for females), whereas Siraro and Shashemene chickens had the lowest weights. SL and BW demonstrated the highest R^2 and F values in both sexes. The greatest coefficients of variation (CV) were observed for WL and BW, ranging from 8.56% to 33.72% in males and from 6.28% to 45.52% in females.

Tables 2, 3 and 4 present descriptive statistics for qualitative traits. The body plumage colour distribution varied significantly across districts for both sexes (Table 2; $p < 0.0001$). Among males, red plumage predominated (52.99%), particularly in Siraro district (90.91%), while white and multi-coloured plumage were uncommon. Females showed different patterns, with brown plumage being most common (50.72%) and Kokima colouration being prevalent in Siraro (41.67%), following brown colour. District-level variation in comb type was more pronounced in females ($p < 0.0001$; Table 3), with the single comb being the most common (44.03%). Rose and pea combs were more common in specific districts.

Most qualitative traits exhibited significant district-specific patterns, except for feather distribution and shank feather presence (Table 4). The majority of chickens had normal feathers (94.52%), though silky feathers were present in Shashemene and Siraro. Yellow shanks predominated (51.69%), especially in Siraro, while white skin was most common overall (62.80%). However, yellow skin prevailed in Shashemene and Siraro. Nearly all chickens had earlobes (97.58%), with white and red or red-spotted white being dominant (36.07%), particularly in Adama, Lume and Bora. Plain heads were most common (66.51%), except in Adama, where snake-like heads predominated (42.31%). While most chickens lacked spurs (55.39%), the majority in Shashemene district possessed spurs.

Table 1. Descriptive statistics and univariate test results for morphometric traits of chickens sampled across six districts (n = 621; 134 males and 487 females). The levels (mean values) not connected by the same letter in a row are significantly different at p ≤ 0.05. Standard deviations (SD) indicate the average variability across all districts. CV, coefficients of variation; BW, body weight; BL, body length; CC, chest circumference; SL, shank length; SC, shank circumference; WS, wingspan; CL, comb length; WL, wattle length; AD, Adama; LM, Lume; BR, Bora; DD, Dodola; SH, Shashemene; SR, Siraro.

Variable	Sex	District						SD	CV, %	R^2	F value	p value
		AD	LM	BR	DD	SH	SR					
BW, g	Male	1412.00^{b}	1854.55^{a}	1256.22^{bc}	1090.48^{cd}	971.43^{d}	977.27^{d}	402.21	31.74	0.58	35.21	< 0.0001
	Female	1206.33^{b}	1435.37^{a}	1085.90^{c}	841.98^{d}	720.99^{e}	715.48^{e}	342.93	34.34	0.61	148.61	< 0.0001
BL, cm	Male	38.52^{b}	40.86^{a}	36.57^{b}	36.62^{b}	36.81^{b}	36.64^{b}	2.79	7.40	0.32	12.14	< 0.0001
	Female	35.79^{a}	36.10^{a}	34.75^{b}	33.16^{c}	33.26^{c}	32.90^{c}	2.32	6.76	0.31	43.57	< 0.0001
CC, cm	Male	27.80^{ab}	29.55^{a}	25.70^{c}	28.86^{ab}	27.33^{bc}	27.09^{bc}	2.63	9.50	0.22	7.40	< 0.0001
	Female	25.44^{c}	26.40^{ab}	24.49^{d}	26.56^{a}	26.23^{abc}	25.67^{bc}	2.08	8.05	0.11	12.25	< 0.0001
SL, cm	Male	9.98^{b}	10.89^{a}	10.09^{b}	8.33^{c}	8.40^{c}	8.25^{c}	1.31	13.95	0.62	42.40	< 0.0001
	Female	8.28^{b}	8.82^{a}	8.22^{b}	6.88^{c}	6.86^{c}	6.88^{c}	1.01	13.23	0.64	171.23	< 0.0001
SC, cm	Male	4.64^{a}	4.98^{a}	4.50^{a}	3.98^{b}	3.92^{b}	3.84^{b}	0.72	16.63	0.34	13.21	< 0.0001
	Female	3.94^{b}	4.16^{a}	3.77^{b}	3.43^{c}	3.30^{cd}	3.16^{d}	0.57	15.68	0.40	63.38	<0.0001
WS, cm	Male	41.36^{b}	45.77^{a}	40.39^{b}	42.43^{b}	41.95^{b}	41.86^{b}	3.62	8.56	0.22	7.11	<0.0001
	Female	38.07^{b}	39.13^{a}	38.09^{b}	37.50^{bc}	36.95^{c}	36.82^{c}	2.37	6.28	0.11	12.06	<0.0001
CL, cm	Male	6.10^{b}	7.68^{a}	6.30^{b}	6.17^{b}	5.67^{b}	5.61^{b}	1.67	26.67	0.18	5.14	0.0002
	Female	3.33^{b}	3.95^{a}	2.95^{cd}	3.09^{bc}	2.99^{bcd}	2.63^{d}	0.91	28.83	0.21	25.16	<0.0001
WL, cm	Male	3.60^{bc}	4.66^{a}	3.37^{bc}	3.93^{ab}	3.10^{bc}	2.84^{c}	1.21	33.72	0.24	7.99	<0.0001
	Female	1.54^{b}	1.85^{a}	1.39^{bc}	1.16^{cd}	1.09^{d}	0.93^{d}	0.61	45.52	0.26	33.79	<0.0001

Table 2. Frequency distribution (percent) of body plumage colour and chi-square test results by sex (n = 621; 134 males and 487 females). Local plumage colour names are as follows: Gebsima, wheaten stripes on black; Teterima, black or red speckles on white; Kokima, white or grey stripes on brown/red; Wesera, mixed white and red; Zigrima, black-and-white spotted (Dana *et al*, 2010; Melesse and Negesse, 2011). AD, Adama; LM, Lume; BR, Bora; DD, Dodola; SH, Shashemene; SR, Siraro.

Colour	District						Total mean	Chi-square	p value
	AD	LM	BR	DD	SH	SR			
Male								69.64	< 0.0001
White	16.00	36.36	13.04	14.29	0.00	0.00	13.43		
Red	32.00	22.73	43.48	61.93	71.43	90.91	52.99		
Gebsima	24.00	9.09	17.39	4.76	14.29	0.00	11.94		
Brown	20.00	13.64	0.00	0.00	0.00	4.55	6.72		
Kokima	0.00	0.00	0.00	4.76	0.00	0.00	0.75		
Wesera	8.00	4.55	21.74	14.29	14.29	4.55	11.19		
Multicolour	0.00	13.64	4.35	0.00	0.00	0.00	2.99		
Female								152.33	< 0.0001
White	12.66	15.66	17.72	9.88	2.47	0.00	9.65		
Black	13.92	7.23	10.13	16.05	0.00	1.19	8.01		
Red	2.53	6.02	12.66	1.23	0.00	0.00	3.71		
Gebsima	2.53	3.61	1.27	1.23	0.00	0.00	1.44		
Teterima	5.06	0.00	3.80	2.47	2.47	0.00	2.26		
Brown	51.90	42.17	44.30	39.51	70.37	55.95	50.72		
Kokima	10.13	16.87	7.59	29.63	23.46	41.67	21.77		
Wesera	0.00	6.02	2.53	0.00	0.00	0.00	1.44		
Zigrima	1.27	2.41	0.00	0.00	0.00	0.00	0.62		
Multicolour	0.00	0.00	0.00	0.00	1.23	1.19	0.41		

Table 3. Frequency distribution (percent) of comb type and chi-square test results by sex (n = 621; 134 males and 487 females). AD, Adama; LM, Lume; BR, Bora; DD, Dodola; SH, Shashemene; SR, Siraro.

Comb type	District						Total mean	Chi-square	p value
	AD	LM	BR	DD	SH	SR			
Male								15.80	0.7292
Single	44.00	45.45	43.48	23.81	33.33	31.82	37.31		
Rose	48.00	50.00	39.13	66.67	47.62	63.64	52.24		
Pea	4.00	0.00	0.00	4.76	4.76	0.00	2.24		
Strawberry	4.00	4.55	13.04	4.76	14.29	4.55	7.46		
Cushion	0.00	0.00	4.35	0.00	0.00	0.00	0.75		
Female								68.53	< 0.0001
Single	31.65	48.19	42.31	49.39	51.85	40.48	44.03		
Rose	30.38	33.73	37.18	23.46	29.63	44.05	33.13		
Pea	21.52	2.41	5.13	22.22	14.81	14.29	13.77		
Walnut	0.00	6.02	2.56	0.00	0.00	0.00	1.44		
Strawberry	15.19	8.43	10.26	4.94	3.70	1.19	7.20		
Cushion	1.27	1.20	2.56	0.00	0.00	0.00	0.82		

Table 4. Frequency distributions (percent) and chi-square test results for qualitative morphological traits (n = 621; 134 males and 487 females). AD, Adama; LM, Lume; BR, Bora; DD, Dodola; SH, Shashemene; SR, Siraro; Ebab-ras, snake-like head; Gutya, crest head.

Trait	District						Total mean	Chi-square	p value
	AD	LM	BR	DD	SH	SR			
Feather morphology								42.38	< 0.0001
Normal	100.00	99.05	100.00	95.10	85.29	87.74	94.52		
Silky	0.00	0.95	0.00	4.90	14.71	12.26	5.48		
Feather distribution								10.84	0.0546
Normal	100.00	97.14	100.00	100.00	100.00	99.06	99.36		
Naked neck	0.00	2.86	0.00	0.00	0.00	0.94	0.64		
Shank colour								127.87	< 0.0001
White	50.96	45.71	39.22	13.73	20.59	9.43	29.95		
Black	8.65	0.95	3.92	2.94	1.96	1.89	3.38		
Yellow	26.92	41.90	37.25	56.86	69.61	77.36	51.69		
Green	1.92	0.00	0.98	0.00	0.00	0.00	0.48		
Grey /grey-blue	11.54	11.43	18.63	26.47	7.84	11.32	14.49		
Shank feather								8.06	0.1529
Absent	96.15	96.19	98.04	98.04	100.00	100.00	98.07		
Present	3.85	3.81	1.96	1.96	0.00	0.00	1.93		
Skin colour								246.07	< 0.0001
White	82.69	82.86	85.29	50.98	30.39	44.34	62.80		
Yellow	12.50	1.90	0.98	47.06	66.67	55.66	30.76		
Blue-black	0.00	0.00	0.00	1.96	2.94	0.00	0.81		
Grey	4.81	15.24	13.73	0.00	0.00	0.00	5.64		
Earlobe presence								14.01	0.0155
Absent	0.96	0.00	0.00	5.88	4.90	2.85	2.42		
Present	99.04	100.00	100.00	94.12	95.10	97.17	97.58		
Earlobe colour								301.28	< 0.0001
White	10.58	5.71	28.43	41.18	25.49	13.21	20.61		
Red	22.12	37.14	18.63	38.24	23.53	28.30	28.02		
Yellow	0.00	0.00	0.00	5.88	27.45	39.62	12.24		
White and red	66.35	56.19	51.96	13.73	14.71	13.21	36.07		
Yellow and red	0.00	0.00	0.00	0.98	8.82	5.66	2.58		
Black	0.96	0.95	0.98	0.00	0.00	0.00	0.48		
Head shape								66.06	< 0.0001
Ebab-ras	42.31	24.76	24.51	23.53	14.71	4.72	22.38		
Gutya	15.38	4.76	16.67	9.80	13.73	6.60	11.11		
Plain/flat	42.31	70.48	58.82	66.67	71.57	88.68	66.51		
Spur presence								32.42	< 0.0001
Absent	55.77	60.95	60.78	59.80	30.39	64.15	55.39		
Present	44.23	39.05	39.22	40.20	69.61	35.85	44.61		

Multivariate analysis

The two-dimensional biplot (Figure 1) illustrates the distribution and associations of qualitative morphological traits across districts. Chickens from Shashemene and Siraro districts clustered closely with yellow skin, while those from Adama, Lume and Bora grouped near grey skin, white plumage, strawberry combs, white–red combs and white shanks. Dodola chickens were strongly associated with plain head and brown plumage. Traits such as silky feathers, yellow–red earlobe, red plumage and yellow skin were positioned far from the origin of the first dimension (Dim 1). Similarly, traits such as Zigirima plumage, cushion comb, Wesera plumage, green shank, silky feather, shank feather, Gebsima plumage, red plumage, naked neck and walnut comb, among others, were distantly located from the origin of Dim 2.

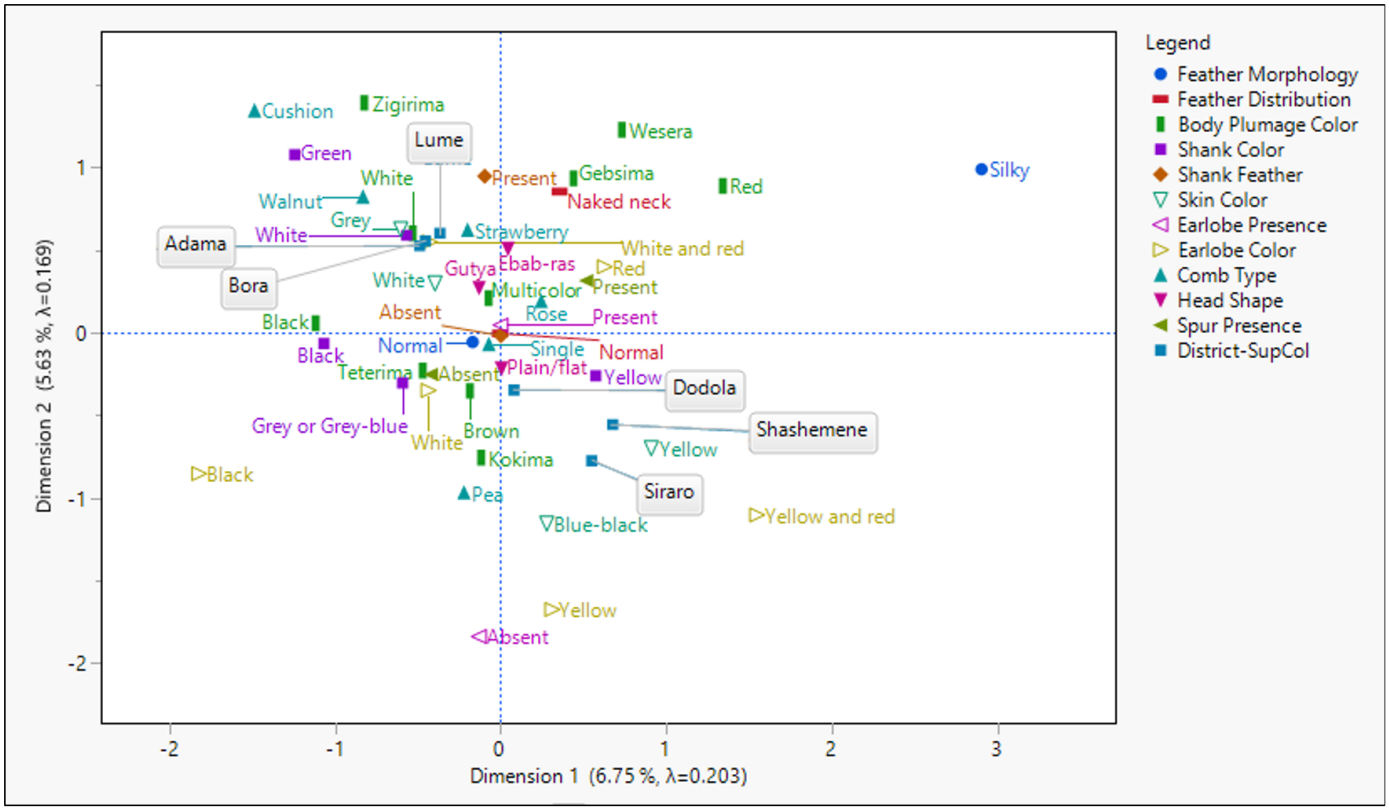


Figure 1. Biplot from correspondence analysis displaying associations between qualitative traits of chickens and their sampling districts.

Table 5 presents the Benzécri-adjusted inertia decomposition, revealing that the first two dimensions collectively explained 70.23% of the total variance (Dim 1: 47.19%; Dim 2: 23.04%). Partial contributions to inertia (Supplemental Table 1) indicate that Dim 1 was primarily defined by silky feathers, red plumage, yellow skin/shank, spur presence, and red earlobes, while Dim 2 was driven by yellow skin/earlobes, red plumage, red-spotted white earlobes, white skin and pea combs.

Table 5. Benzécri-adjusted inertia decomposition of qualitative morphological traits (n = 621; 134 males and 487 females).

Dimension	Inertia	Adjusted inertia	Percent	Cumulative percent
1	0.2025	0.0151	47.19	47.19
2	0.1689	0.0074	23.04	70.23
3	0.1513	0.0044	13.76	84.03
4	0.1298	0.0018	5.74	89.76
5	0.1227	0.0012	3.83	93.60
6	0.1150	0.0007	2.20	95.79
7	0.1111	0.0005	1.55	97.34
8	0.1074	0.0003	1.02	98.36
9	0.1040	0.0002	0.65	99.01
10	0.1018	0.0001	0.45	99.46
11	0.0998	0.0001	0.30	99.76
12	0.0967	0.0000	0.13	99.89
13	0.0954	0.0000	0.08	99.96
14	0.0938	0.0000	0.03	100.00
15	0.0920	0.0000	0.00	100.00

In this study, QDA was employed instead of linear discriminant analysis (LDA) due to the presence of unequal covariance matrices across the different chicken populations. The analysis successfully classified chickens into their respective districts of origin (Table 6). Chickens from Bora (88.31% of females and 91.30% of males) and Lume (67.10% of females and 90.90% of males) showed the highest classification accuracy, whereas Adama females and Dodola males had the lowest. Misclassification patterns revealed that 36.71% of female and 24.00% of male chickens from Adama were misassigned to Bora, while 19.00% of females and 16% of males were misassigned to Lume. Similarly, 24.39% of females and 9.09% of males from Lume were misclassified to Bora. Additionally, 22.10% of female and 23.80% of male chickens from Dodola were erroneously grouped into Shashemene, with a similar proportion of both sexes misassigned to Siraro. Furthermore, a substantial proportion of female and male chickens from Shashemene and Siraro were reciprocally misclassified into each other's districts.

The SDA identified seven morphometric traits (SL, BW, CC, CL, WS, SC and BL) for females and only four traits (SL, CC, BW and WS) for males as the most discriminative variables (Table 7). The analysis revealed that BW, CC and SL in females and SL in males exhibited the highest discriminatory power (indicated by their higher F and Wilks' Lambda values; $p < 0.0001$), while BL in females and WS in males were less effective for group discrimination (reflected by their lower F values).

Table 6. Classification results showing the number (and percentage) of correctly assigned male and female chickens (n = 621; 134 males and 487 females). AD, Adama; LM, Lume; BR, Bora; DD, Dodola; SH, Shashemene; SR, Siraro.

District	AD	LM	BR	DD	SH	SR	Total
Female							
AD	33 (41.80)	15 (19.00)	29 (36.71)	0 (0.00)	2 (2.53)	0 (0.00)	79 (100.00)
LM	6 (7.32)	55 (67.10)	20 (24.39)	1 (1.22)	0 (0.00)	0 (0.00)	82 (100.00)
BR	2 (2.60)	6 (7.79)	68 (88.31)	0 (0.00)	0 (0.00)	1 (1.30)	77 (100.00)
DD	1 (1.30)	0 (0.00)	2 (2.60)	41 (53.20)	17 (22.10)	16 (20.80)	77 (100.00)
SH	2 (2.53)	0 (0.00)	1 (1.27)	7 (8.86)	48 (60.80)	21 (26.60)	79 (100.00)
SR	1 (1.25)	0 (0.00)	4 (5.00)	4 (5.00)	24 (30.00)	47 (58.80)	80 (100.00)
Male							
AD	13 (52.00)	4 (16.00)	6 (24.00)	0 (0.00)	2 (8.00)	0 (0.00)	25 (100.00)
LM	0 (0.00)	20 (90.90)	2 (9.09)	0 (0.00)	0 (0.00)	0 (0.00)	22 (100.00)
BR	1 (4.35)	0 (0.00)	21 (91.30)	1 (4.35)	0 (0.00)	0 (0.00)	23 (100.00)
DD	0 (0.00)	1 (4.76)	0 (0.00)	10 (47.60)	5 (23.80)	5 (23.80)	21 (100.00)
SH	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	16 (84.20)	3 (15.80)	19 (100.00)
SR	0 (0.00)	0 (0.00)	0 (0.00)	3 (13.60)	3 (13.60)	16 (72.70)	22 (100.00)

Table 7. Stepwise selection summary of the most discriminative morphometric traits for female and male chickens (n = 621; 134 males and 487 females). BW, body weight; BL, body length; CC, chest circumference; SL, shank length; SC, shank circumference; WS, wingspan; CL, comb length.

Step	Entered	F value	p value	Wilks' Lambda
Female				
1	SL	19.73	< 0.0001	0.36
2	BW	33.47	< 0.0001	0.28
3	CC	23.33	< 0.0001	0.20
4	CL	4.14	0.0011	0.16
5	WS	6.88	< 0.0001	0.19
6	SC	4.40	0.0006	0.17
7	BL	3.27	0.0065	0.18
Male				
1	SL	14.45	< 0.0001	0.38
2	CC	5.28	0.0002	0.26
3	BW	7.02	< 0.0001	0.19
4	WS	5.09	0.0003	0.15

The CDA extracted five and four canonical discriminant functions (CAN) in females and males, respectively (Table 8). The first two functions explained approximately 98% of the variance in both sexes. CAN1 accounted for the majority of the variance (91.77% for females, 81.58% for males), while CAN2 explained 6.09% and 16.55% for females and males, respectively. CAN1 exhibited greater correlations (0.8857 and 0.8596 for females and males, respectively) with the groups. For both sexes, CAN1 had a lower Wilks' Lambda of 0.16.

Figure 2 shows the relationships between the morphometric traits selected through SDA and the two most discriminating functions across the six districts. For female chickens, CAN1 effectively discriminated between chickens from Adama, Lume and Bora populations, while CAN2 discriminated among Lume, Dodola and Shashemene populations. For male chickens, CAN1 differentiated chickens from Adama, Lume and Bora, whereas CAN2 distinguished between Lume and Dodola populations. In females, CAN1 was strongly correlated with SL, BW, SC and BL, while CAN2 was predominantly associated with CC and CL, as evidenced by higher loadings (Table 9). Male chickens exhibited different patterns, with CAN1 being most strongly linked to SL and BW, while CAN2 was correlated with CC, WS and BW. In the biplot, districts from East Shoa were positioned on the right (high CAN1) for both sexes, with Lume exhibiting the highest CAN1 value, while districts from West Arsi clustered on the left (low CAN1).

Table 8. Summary of canonical discriminant functions for male and female chickens (n = 621; 134 males and 487 females). CAN, canonical function.

CAN	Eigenvalue	Percent (%)	Cumulative %	Canonical corr.	Wilks' Lambda	F value	p value
Female							
CAN1	3.6409	91.77	91.77	0.8857	0.1597	30.39	<0.0001
CAN2	0.2417	6.09	97.86	0.4412	0.7412	6.04	<0.0001
CAN3	0.0551	1.39	99.25	0.2285	0.9203	2.61	0.0007
CAN4	0.0198	0.50	99.75	0.1393	0.9710	1.72	0.0895
CAN5	0.0098	0.25	100.00	0.0988	0.9902	1.53	0.2060
Male							
CAN1	2.8296	81.58	81.58	0.8596	0.1558	15.37	<0.0001
CAN2	0.5742	16.55	98.13	0.6039	0.5965	5.90	<0.0001
CAN3	0.0593	1.71	99.84	0.2366	0.9390	1.33	0.2428
CAN4	0.0054	0.16	100.00	0.0733	0.9946	0.34	0.7122

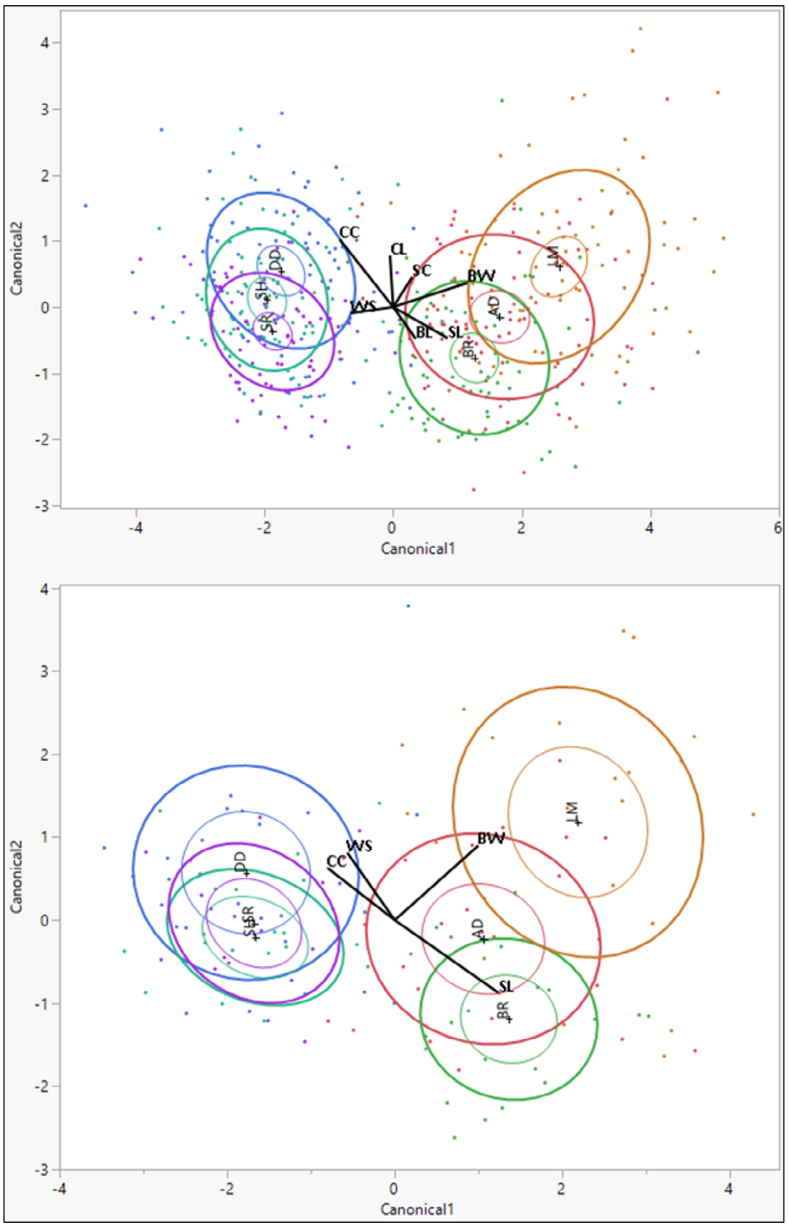


Figure 2. Biplot of canonical discriminant functions and structure loadings for female (upper) and male (lower) chickens. BW, body weight; BL, body length; CC, chest circumference; SL, shank length; SC, shank circumference; WS, wingspan; CL, comb length; AD, Adama; LM, Lume; BR, Bora; DD, Dodola; SH, Shashemene; SR, Siraro.

Table 9. Standardized canonical coefficients and canonical structures for female and male chickens from six districts (n = 621; 134 males and 487 females). CAN, canonical function.

Trait	Standardized canonical coefficient		Canonical structure	
	CAN 1	CAN2	CAN 1	CAN2
Female				
SL	0.5636	–0.3135	0.9031	0.0122
BW	0.7737	0.2445	0.8587	0.3827
CC	–0.5539	0.6804	–0.1274	0.7539
CL	–0.0323	0.5191	0.3837	0.6564
WS	–0.4422	–0.0626	0.3564	0.2370
SC	0.1965	0.3021	0.6971	0.2761
BL	0.2330	–0.3188	0.6172	0.1335
Male				
SL	0.8242	–0.5826	0.9177	0.1345
CC	–0.5370	0.4198	0.0742	0.7846
BW	0.6652	0.5986	0.7964	0.6023
WS	–0.3777	0.5410	0.1688	0.7123

Analysis of interclass distances (Table 10) based on the differences between class means on discriminant functions for CAN1, which accounted for the majority of the variance in both sexes, revealed clear geographical differentiation. Lume district chickens were significantly different from those in all three districts in the West Arsi zone (Shashemene, Siraro and Dodola, in that order). CAN1 also effectively discriminated between Adama and the three districts in West Arsi, as well as between Bora and those same districts. Notably, districts within the same geographical zone clustered more closely together.

Table 10. The distance between mean vectors of each class for male (above diagonal) and female (below diagonal) chickens (n = 621; 134 males and 487 females). AD, Adama; LM, Lume; BR, Bora; DD, Dodola; SH, Shashemene; SR, Siraro.

District		AD	LM	BR	DD	SH	SR
					Male		
AD			1.13	0.30	2.83	2.72	2.74
LM		0.92		0.82	3.96	3.85	3.86
BR	Female	0.39	1.32		3.14	3.03	3.04
DD		3.40	4.32	3.01		0.11	0.10
SH		3.62	4.54	3.22	0.22		0.02
SR		3.53	4.45	3.14	0.13	0.09	

Discussion

Univariate analysis

In animal breeding, metrics such as body weight, length, and wither height serve as valuable proxies for economic performance due to their strong correlation with production output (FAO, 2012). The highly significant differences in morphometric traits among districts observed in this study align with previous findings in Ethiopia (Getachew *et al*, 2016; Assefa and Melesse, 2018; Tareke *et al*, 2018; Melesse *et al*, 2021; Mekonnen *et al*, 2023; Belay *et al*, 2024; Chebo *et al*, 2024) and other countries (Habimana *et al*, 2021; Liswaniso *et al*, 2024; Yaemkong *et al*, 2024). Variations in specific measurements across studies are likely attributable to differing management systems, as noted by Melesse *et al* (2021). These authors further emphasized that phenotypic variations may arise from genotypic and environmental factors and their interactions, which are crucial for long-term genetic adaptation to specific production environments.

Live BW is an economically important trait in animal production. In this study, chickens from the Lume district were the heaviest, followed by those from Adama. These values exceed those reported from other parts of Ethiopia (Hassen *et al*, 2007; Getu *et al*, 2014; Nigussie *et al*, 2015; Bekele *et al*, 2021) and other countries (Daikwo *et al*, 2011; Rotimi *et al*, 2016) but are lower than reports from elsewhere (Maharani *et al*, 2021; Moto and Rubanza, 2023). The BW of male (1854.55) and female (1435.37) chickens from Lume were higher than those found by Aklilu *et al* (2013) and Tamirat *et al* (2023). Such geographic variations are likely associated with differences in production environment, management practices, genetic background and the age at measurement.

Similarly, chickens from Lume and Adama were longer-bodied. The BL of male chickens from Adama was similar to that of males from other districts. Chickens from these two districts were longer than those reported from other parts of Ethiopia (Getu *et al*, 2014; Nigussie *et al*, 2015; Tamirat *et al*, 2023; Markos *et al*, 2024) and Nigeria (Daikwo *et al*, 2011), but shorter than chickens in other Ethiopian studies (Balcha *et al*, 2022) and other countries (Rotimi *et al*, 2016; Habimana *et al*, 2021; Moto and Rubanza, 2023). CC was highest in male chickens from Lume and in female chickens from Dodola. These values were higher than those in several previous reports (Nigussie *et al*, 2015; Balcha *et al*, 2022; Bayou *et al*, 2022; Mekonnen *et al*, 2023; Belay *et al*, 2024), though higher values have been documented in other Ethiopian studies (Aklilu *et al*, 2013; Tadese *et al*, 2024) and other country (Moto and Rubanza, 2023).

Higher values for shank traits (SL and SC) were recorded in chickens from Lume, followed by Adama and Bora. The SL values in these districts exceeded those from previous studies (Daikwo *et al*, 2011; Getu *et al*, 2014; Nigussie *et al*, 2015; Habimana *et al*, 2021; Maharani *et al*, 2021; Mekonnen *et al*, 2023), though longer shanks have also been reported (Aklilu *et al*, 2013; Markos *et al*, 2024). Long shanks are an adaptation for heat dissipation in tropical climates (Aklilu *et al*, 2013), which may explain their prevalence in the districts within Ethiopia's Central Rift Valley region. Udeh *et al* (2011) reported a relationship between SL and semen traits that varies by breed. They found SL to be a significant predictor of sperm motility in exotic cocks and the proportion of live sperm in local cocks. Long shanks in indigenous chickens, which are suitable for fast running, may also facilitate predator avoidance (Besbes, 2009; Ngeno *et al*, 2014).

Lume chickens exhibited greater WS, with values in other districts being closely related. Chickens from other parts of Ethiopia (Nigussie *et al*, 2015; Bekele *et al*, 2021; Bayou *et al*, 2022) and Rwanda (Habimana *et al*, 2021) exhibited greater WS, while lower values have been reported (Markos *et al*, 2024; Tadese *et al*, 2024). Udeh *et al* (2011) found a strong correlation between wing length and the proportion of live sperm in local cocks (r = 0.59), indicating its utility as a

predictor for this trait.

Chickens from Lume had longer combs than those from other districts, exceeding lengths reported previously (Nigussie *et al*, 2015; Bekele *et al*, 2021; Belay *et al*, 2024; Markos *et al*, 2024; Tadese *et al*, 2024). The relationship between comb size and semen quality is complex and varies by breed. While inverse correlations have been observed in some studies (Udeh *et al*, 2011; Navara *et al*, 2012), Udeh *et al* (2011) also found CL to be a significant positive predictor of sperm concentration in exotic cocks. A relationship between comb size and BW has also been reported (Suyatno *et al*, 2023). Chickens from Lume had longer wattles than those in other districts, with values higher than those in previous reports (Nigussie *et al*, 2015; Bekele *et al*, 2021; Habimana *et al*, 2021; Maharani *et al*, 2021; Belay *et al*, 2024; Tadese *et al*, 2024). However, longer wattles have been found in other Ethiopian regions (Markos *et al*, 2024).

This study clearly demonstrates that chickens from Lume district consistently exhibited the highest values for key morphometric traits, including BW, identifying them as a high-priority population for selective breeding. Notably, Lume is a district where small-scale exotic chicken farming is common. Smallholder farmers often keep these breeds under the assumption that they are high-yielding and can improve indigenous stock through crossbreeding. Although farmers who owned exotic chickens were purposely excluded from this study, their uncontrolled distribution likely leads to introgression with indigenous populations. Therefore, the superior morphometric values in Lume may be associated with crossbreeding rather than representing the pure potential of the indigenous gene pool.

Significant sexual dimorphism was observed in all morphometric measurements, with males being larger than females, consistent with global studies (Liyanage *et al*, 2015; Mekonnen *et al*, 2023; Yaemkong *et al*, 2024; Begna *et al*, 2025). This dimorphism likely results from a combination of factors, including hormonally mediated superior muscle development in males (Semakula *et al*, 2011), evolutionary selective pressures such as intra-sexual competition or divergent parental care (Owens and Hartley, 1998), and different growth rates. This difference must be considered when designing breeding programmes for meat improvement (Habimana *et al*, 2021).

These morphometric traits, being polygenic, exhibit continuous variation (FAO, 2012). The highly significant ($p < 0.001$) differences for all traits confirm their strong discriminatory power. SL and BW demonstrated the highest F-values, revealing them as the most distinct traits and making them excellent for differentiating the groups. Conversely, WS and CC showed the weakest differentiation, making them poor for distinguishing among these populations. Furthermore, SL ($R^2 = 0.64$ in females, 0.62 in males) and BW ($R^2 = 0.61$ in females, 0.58 in males) had the greatest model explanatory power, meaning that district-level differences account for most of their variation. This aligns with Getachew *et al* (2016), who reported high explanatory power for BW. The observed CV across districts (7.40–33.72% in males; 6.28–45.52% in females) indicates substantial phenotypic heterogeneity, with WL and BW contributing the most, consistent with prior findings (Getachew *et al*, 2016).

Qualitative morphological traits such as plumage colour, comb type and skin pigmentation are markers of genetic diversity and adaptive features. They also hold economic value by influencing market preferences, product differentiation and breed valuation (Dana *et al*, 2010). Studying this variation provides evolutionary insights and reflects the socioeconomic context of subsistence farmers (Desta *et al*, 2013; Yussif *et al*, 2023). While these traits have less direct impact on production than quantitative traits, their adaptive significance makes them relevant for conservation (FAO, 2012).

This study revealed significant geographic variation in qualitative traits, confirmed by χ^2 tests. This indicates that the distribution of these traits is not the same across the districts. Higher χ^2 values for traits like earlobe colour, skin colour, feather colour and shank colour indicate their distribution is non-random across districts. This distinct geographic patterning, supported by MCA, suggests the existence of locally adapted populations, providing a phenotypic basis for targeted conservation and breeding strategies.

Avian colouration is influenced by complex interactions between pigments (melanins, carotenoids, psittacofulvins) and structural colours (Price-Waldman and Stoddard, 2021). Indigenous chickens show greater plumage diversity than standardized breeds (Tixier-Boichard *et al*, 2008). As also noted by various authors (Besbes, 2009; Bibi *et al*, 2021; Lawal and Hanotte, 2021; Chebo *et al*, 2023), indigenous chickens in this study are invariably coloured birds that display various plumage colours across the districts. This variation results from feather development mechanisms, genetics, and human selection (Cabarles *et al*, 2012).

The prevalence of red plumage in males and brown plumage in females is also common in different Ethiopian (Dana *et al*, 2010; Bekele *et al*, 2021; Begna *et al*, 2025) and Philippine (Cabarles *et al*, 2012) indigenous chickens. This pattern, inherited from the red junglefowl and reinforced by natural selection, provides camouflage (Besbes, 2009; Cabarles *et al*, 2012). However, a discrete choice experiment in rural Ethiopia found a farmer preference for white plumage over red (Terfa *et al*, 2019).

Regardless of sex differences, brown plumage is dominant in Nigeria (Daikwo *et al*, 2011), Ethiopia (Moreda *et al*, 2014), and Zambia (Liswaniso *et al*, 2024), while red is most common in other Ethiopian studies (Melesse and Negesse, 2011; Nigussie *et al*, 2015; Balcha *et al*, 2022). In the Borena zone of Ethiopia, the majority of both male and female chickens have white plumage, followed by red and brown plumage (Wario *et al*, 2021), whereas most chickens from North Wollo are white, followed by black and red (Achenef *et al*, 2023). Black plumage, rare or absent here, is dominant in Algeria (Moula *et al*, 2012), Nigeria (Egahi *et al*, 2010; Ige *et al*, 2012), Thailand (Buranawit *et al*, 2016) and Sri Lanka (Liyanage *et al*, 2015). Multicoloured plumage is most common in Kenya (Otecko *et al*, 2019), Tanzania (Moto and Rubanza, 2023) and Indonesia (Maharani *et al*, 2021), while a greyish mixture is predominant in Pakistan (Bibi *et al*, 2021).

Most chickens in this study had normal feather morphology and distribution, consistent with reports from Ethiopia (Melesse and Negesse, 2011; Nigussie *et al*, 2015; Assefa and Melesse, 2018; Mustefa *et al*, 2021; Wario *et al*, 2021; Balcha *et al*, 2022; Bayou *et al*, 2022; Achenef *et al*, 2023; Chebo *et al*, 2023; Muluneh *et al*, 2023; Tamirat *et al*, 2023; Tadese *et al*, 2024), Rwanda (Habimana *et al*, 2021), Ghana (Birteeb and Boakye, 2020) and Uganda (Yussif *et al*, 2023). While Dana *et al* (2010) reported nearly equal proportions

of normal (58%) and silky (42%) feather types, most studies (Ige *et al*, 2012; Rotimi *et al*, 2016; Begna *et al*, 2025) confirm normal feathering is predominant.

The very low proportions or complete absence of silky, naked neck, and frizzle feather phenotypes in the current study suggest that farmers actively avoid selecting for these traits, potentially due to cultural stigma, as in Nigeria (Ige *et al*, 2012; Rotimi *et al*, 2016). Despite this, these genotypes are valuable for disease resistance (Ngeno *et al*, 2014), thermotolerance, improved feed-to-meat conversion and growth performance (Duguma, 2006; Dana *et al*, 2010; Melesse and Negesse, 2011). Naked-neck chickens are also known for their improved immune competence, better meat quality (e.g. lower cholesterol, higher dressing percentage), and easier processing due to less feather plucking (Desta, 2021). About 10 to 12% of naked-neck chickens have been reported in southwestern Ethiopia (Assefa and Melesse, 2018; Bayou *et al*, 2022).

In Ethiopia, farmers have historically culled single-combed chickens, reducing their frequency (Desta *et al*, 2013; Muluneh *et al*, 2023). In this study, single combs were most common in females and rose combs in males, consistent with Negassa *et al* (2014). This contrasts with Chebo *et al* (2023), who found single combs most common in males, with females having rose and single combs. Despite a reported fertility reduction in roosters homozygous for the rose comb allele (McLean and Froman, 1996), their high prevalence here may be driven by cultural and market preferences (Chebo *et al*, 2023). The high prevalence of single combs may also relate to heat dissipation (Moreda *et al*, 2014; Rotimi *et al*, 2016). Selection practices should consider the association of comb types with production and fertility (Chebo *et al*, 2023).

A single comb was the predominant type in both sexes of indigenous chickens in Ethiopia (Mustefa *et al*, 2021; Wario *et al*, 2021; Begna *et al*, 2025), Uganda (Yussif *et al*, 2023), Zambia (Liswaniso *et al*, 2024), the Philippines (Cabarles *et al*, 2012; Picardal *et al*, 2015), Sri Lanka (Liyanage *et al*, 2015), Bangladesh (Sarker *et al*, 2014) and Indonesia (Maharani *et al*, 2021). Regardless of sex differences, this comb type has been recorded in the majority of indigenous chickens in Nigeria (Egahi *et al*, 2010; Daikwo *et al*, 2011; Ige *et al*, 2012; Rotimi *et al*, 2016), Rwanda (Habimana *et al*, 2021), Ghana (Birteeb and Boakye, 2020), Kenya (Otecko *et al*, 2019), Tanzania (Moto and Rubanza, 2023), Ethiopia (Melesse and Negesse, 2011; Moreda *et al*, 2014; Assefa and Melesse, 2018; Alebachew *et al*, 2019; Bekele *et al*, 2021; Balcha *et al*, 2022; Bayou *et al*, 2022; Achenef *et al*, 2023; Tamirat *et al*, 2023), Thailand (Buranawit *et al*, 2016) and Pakistan (Bibi *et al*, 2021). However, pea combs occur in relatively high proportions in some Ethiopian studies, typically followed by single and rose combs (Hassen *et al*, 2007; Dana *et al*, 2010; Getachew *et al*, 2016). It is also the second most prevalent type, following rose combs (Nigussie *et al*, 2015).

A high proportion of yellow shanks was recorded, consistent with findings in Ethiopia (Hassen *et al*, 2007; Dana *et al*, 2010; Melesse and Negesse, 2011; Desta *et al*, 2013; Negassa *et al*, 2014; Nigussie *et al*, 2015; Assefa and Melesse, 2018; Bekele *et al*, 2021; Mustefa *et al*, 2021; Balcha *et al*, 2022; Bayou *et al*, 2022; Chebo *et al*, 2023; Muluneh *et al*, 2023; Tamirat *et al*, 2023), Uganda (Yussif *et al*, 2023), Nigeria (Daikwo *et al*, 2011), the Philippines (Cabarles *et al*, 2012), Sri Lanka (Liyanage *et al*, 2015), Rwanda (Habimana *et al*, 2021), Tanzania (Moto and Rubanza, 2023), Pakistan (Bibi *et al*, 2021) and Indonesia (Maharani *et al*, 2021). This prevalence is likely associated with farmer preferences for this phenotype (Desta *et al*, 2013). However, white shanks are more common in Zambia (Liswaniso *et al*, 2024), Ghana (Birteeb and Boakye, 2020), Nigeria (Rotimi *et al*, 2016) and Ethiopia (Wario *et al*, 2021). Both white and yellow shanks are predominantly found in Bangladesh (Sarker *et al*, 2014) and Ethiopia (Achenef *et al*, 2023), which likely suggests adaptations to heat dissipation in tropical climates (Moreda *et al*, 2014). Black shanks are predominant in Nigeria (Egahi *et al*, 2010) and Thailand (Buranawit *et al*, 2016). In domestic chickens, shank colour variation is caused by a combination of genes affecting carotenoid and melanin pigmentations, polygenic modifiers and environmental factors (Chebo *et al*, 2023).

Nearly all chickens from the six districts had featherless shanks, consistent with other Ethiopian (Hassen *et al*, 2007; Moreda *et al*, 2014; Bekele *et al*, 2021; Chebo *et al*, 2023; Muluneh *et al*, 2023) and Nigerian (Egahi *et al*, 2010) reports. However, equal proportions of chickens in Ethiopia have shanks with and without feathers (Begna *et al*, 2025). While feathered shanks are often considered an adaptation to colder climates (Ngeno *et al*, 2014), their absence in the cooler highland district of Dodola suggests other influencing factors.

Skin colour variation results from hybridization, inheritance, mutations in pigmentation genes (Cabarles *et al*, 2012), and dietary carotenoids (Dana *et al*, 2010). Consistent with the present findings, white skin is dominant in several Ethiopian regions (Nigussie *et al*, 2015; Alebachew *et al*, 2019; Balcha *et al*, 2022; Achenef *et al*, 2023; Muluneh *et al*, 2023; Begna *et al*, 2025) and other countries, including Zambia (Liswaniso *et al*, 2024), Ghana (Birteeb and Boakye, 2020), and the Philippines (Cabarles *et al*, 2012; Picardal *et al*, 2015), often followed by yellow. Yellow skin is predominant in some Ethiopian studies (Negassa *et al*, 2014; Assefa and Melesse, 2018; Tamirat *et al*, 2023) and is common in commercial stocks (Eriksson *et al*, 2008). White and yellow skin colours are also common in Pakistan (Bibi *et al*, 2021). Red was identified as the second most common skin colour after white in Ethiopia (Wario *et al*, 2021). As birds cannot synthesize carotenoids (Price-Waldman and Stoddard, 2021), differences in skin colour across indigenous chicken populations are likely driven by variations in local scavenging feed resources.

The current findings align with those of previous studies from Ethiopia (Moreda *et al*, 2014; Nigussie *et al*, 2015; Getachew *et al*, 2016; Alebachew *et al*, 2019; Chebo *et al*, 2023; Muluneh *et al*, 2023) and the Philippines (Picardal *et al*, 2015), where white-red was the most prevalent colour, followed by red and white. However, Dana *et al* (2010) reported a different pattern, with red being the most common colour, followed by white. A more distinct exception was reported by Bekele *et al* (2021), who identified yellow as the predominant colour, followed by red and white–red. Furthermore, a majority of Thai native chickens were reported to have black earlobes (Buranawit *et al*, 2016). Red earlobes have been identified as the dominant type in indigenous chickens from multiple countries, including Ethiopia (Melesse and Negesse, 2011; Desta *et al*, 2013; Negassa *et al*, 2014; Assefa and Melesse, 2018; Mustefa *et al*, 2021; Balcha *et al*, 2022; Bayou *et al*, 2022; Achenef *et al*, 2023; Tamirat *et al*, 2023; Tadese *et al*, 2024; Begna *et al*, 2025), Rwanda (Habimana *et al*,

2021), Zambia (Liswaniso *et al*, 2024), Indonesia (Maharani *et al*, 2021), and Pakistan (Bibi *et al*, 2021). Conversely, white earlobes were reported as the dominant phenotype in studies from Ethiopia (Duguma, 2006), Ghana (Birteeb and Boakye, 2020), and Nigeria (Egahi *et al*, 2010; Ige *et al*, 2012).

Consistent with observations in indigenous chickens from various Ethiopian regions (Dana *et al*, 2010; Moreda *et al*, 2014; Alebachew *et al*, 2019; Bekele *et al*, 2021; Mustefa *et al*, 2021; Wario *et al*, 2021; Balcha *et al*, 2022; Bayou *et al*, 2022; Tamirat *et al*, 2023; Begna *et al*, 2025), Nigeria (Egahi *et al*, 2010; Rotimi *et al*, 2016) and Ghana (Birteeb and Boakye, 2020), the plain head type was predominant among chickens in the present study. In contrast, snake-like heads are more common than other head shapes in chickens from certain Ethiopian regions (Negassa *et al*, 2014; Achenef *et al*, 2023). While spurs were absent in most chickens in this study, regardless of sex, Mustefa *et al* (2021) reported that spurs were present in most males but absent in most females, suggesting sex-dependent expression of this trait.

As noted by Desta *et al* (2013), the significant variation across locations reflects the combined influence of ecological factors (climate, geography) and traditional breeding history. This broad diversity signifies a substantial gene pool for genetic improvement through selective breeding and provides a foundation for designing effective conservation strategies for sustainable utilization of indigenous chicken genetic resources (Tixier-Boichard *et al*, 2008; Dessie *et al*, 2011).

Multivariate analysis

While univariate analysis offers simplicity by examining individual variables in isolation, multivariate analysis provides more comprehensive and biologically meaningful insights by evaluating multiple interrelated variables simultaneously. MCA is an exploratory technique specifically designed to analyze relationships among categorical variables (FAO, 2012). This approach generates a low-dimensional graphical map of variable categories, revealing clustering patterns and associations that pairwise tests (e.g. chi-square) cannot detect (Abdi and Valentin, 2007; Sourial *et al*, 2010; Fithian and Josse, 2017).

The resulting biplot (Figure 1) clearly demonstrates these relationships, showing strong associations between morphological variables and their geographic origins, a pattern also observed in other studies (Nigussie *et al*, 2015; Chebo *et al*, 2023). For instance, the clustering of yellow skin near the Shashemene and Siraro districts indicates a strong association, reflecting the high observed frequency of this trait in both locations (Table 4). The proximity of other qualitative traits to their respective sampling districts suggests meaningful biological–geographic relationships. In MCA, dimensions represent the largest deviations from variable independence (Sourial *et al*, 2010). Consequently, traits positioned farther from the origin, such as silky feathers (Dim 1) and Zigirima plumage colour (Dim 2), contribute most significantly to the observed patterns. A variable's distance from the origin corresponds to its power to differentiate populations (Chebo *et al*, 2023).

A recognized limitation of MCA is its tendency to underestimate the total explained variance (inertia), often requiring adjustments (Benzécri or Greenacre corrections) for accurate interpretation (Abdi and Valentin, 2007; Sourial *et al*, 2010; Camiz and Gomes, 2016; Khangar and Kamalja, 2017). Compared to Greenacre's conservative approach, the Benzécri method applies a more robust eigenvalue reweighting (Veflen *et al*, 2017). Given that the principal inertias of a Burt matrix produce numerous small eigenvalues, this study utilized Benzécri-adjusted inertias for more reliable variance estimation.

The first two dimensions in this study explained a greater proportion of the total variance than some previous reports. For instance, Chebo *et al* (2023) reported that the first two dimensions collectively explained 20.21%, 15.73%, 30.59% and 32.00% of the variance for different trait groups. Similarly, Nigussie *et al* (2015) and Belay *et al* (2024) reported 29.85% and 14.48%, respectively. In contrast, a study on Guinea fowl reported a higher value of 89.69% (Traoré *et al*, 2018), compared to the 70.23% obtained here.

Partial contributions to inertia quantify how much each category contributes to the variance explained by each dimension. Higher values indicate a more significant role in defining that axis. For example, in this study (Supplemental Table 1), the distinction between silky and non-silky feathers was the primary contributor to Dimension 1, while the presence or absence of yellow earlobes was more influential in Dimension 2. These contributions help identify the key traits driving population variation.

The MCA revealed significant geographic patterning, distinctly separating chickens from the East Shoa zone (Adama, Lume, Bora) and the West Arsi zone (Shashemene, Siraro), characterized by traits like grey skin and white plumage versus yellow skin, respectively. Despite its utility, MCA has been used in only a limited number of studies to characterize indigenous chickens in Ethiopia (Nigussie *et al*, 2015; Chebo *et al*, 2023; Muluneh *et al*, 2023; Belay *et al*, 2024), and none prior to this study in the present research area. The findings in this study can help to address this limitation. These findings can support the establishment of location-specific conservation programmes to maintain unique traits as distinct genetic resources.

Discriminant analysis is another multivariate technique that uses quantitative predictors (morphometric variables) to differentiate among categorical groups (districts). The QDA employed here classified chickens into their districts of origin with accuracies ranging from 41.80% to 88.31% for females and 47.60% to 91.30% for males. These correct classification rates are consistent with some previous studies (Picardal *et al*, 2015; Tareke *et al*, 2018; Mustefa *et al*, 2021; Muluneh *et al*, 2023; Chebo *et al*, 2024). However, it is lower than others that reported accuracies exceeding 80%, sometimes reaching 100% (Aklilu *et al*, 2013; Daikwo *et al*, 2015; Getachew *et al*, 2016; Yakubu and Ari, 2018; Melesse *et al*, 2021; Markos *et al*, 2024). Similarly, Kefelegn *et al* (2016) correctly classified 74% to 92% of chickens (both sexes) into their respective sampling locations, while Wario *et al* (2021) achieved 75% to 78% accuracy for females and 93% to 96% accuracy for males.

The higher correct classification rates for Bora and Lume suggest greater phenotypic homogeneity within chickens of these districts, but distinctness from other populations. In contrast, lower classification rates in the remaining districts indicate higher internal diversity. The overall classification success demonstrates homogeneity within populations relative to the variation between them (Melesse *et al*, 2021). Such distinctness among different geographic populations

suggests location-specific conservation and breeding strategies. The misclassification observed in Adama, Dodola, Shashemene and Siraro may be attributed to overlapping morphometric traits or insufficiently distinct discriminant features, a phenomenon also noted in previous findings (Tareke *et al*, 2018; Muluneh *et al*, 2023).

SDA identified seven key discriminating traits for females and four for males. The number of traits selected varies across studies, suggesting that morphometric differentiation is context-dependent and influenced by genetic background, environment and management practices (Ajayi *et al*, 2012; Daikwo *et al*, 2015; Getachew *et al*, 2016; Kefelegn *et al*, 2016; Tareke *et al*, 2018; Mustefa *et al*, 2021; Wario *et al*, 2021; Bekele *et al*, 2022; Muluneh *et al*, 2023; Tadese *et al*, 2024). This highlights the importance of population-specific trait selection for characterization and conservation. Some of the discriminating variables identified here, such as SL and BW, align with previous findings (Ajayi *et al*, 2012; Aklilu *et al*, 2013; Bekele *et al*, 2022).

CDA constructs CAN functions that maximize separation among groups. While the number of significant functions varies across studies, typically either the first function or the first two functions explain most of the variance. In this study, the first two CAN explained 98% of the variance in both sexes, with CAN1 alone accounting for 91.77% of the variance in females and 81.58% of the variance in males. This high explanatory power is consistent with those of Chebo *et al* (2024), who reported that the first two functions captured 95.6% of the between-population variability. Similarly, other studies (Getachew *et al*, 2016; Kefelegn *et al*, 2016; Mustefa *et al*, 2021; Belay *et al*, 2024) reported that CAN1 alone explained 80 to 100% of the variance in both sexes.

However, some studies reported a lower variance explained by the first two functions. For instance, Tareke *et al* (2018) and Melesse *et al* (2021) reported that these functions accounted for 71.5% and 88% of the variance, respectively, regardless of sex. Muluneh *et al* (2023) reported 82% (females) and 68% (males), while Markos *et al* (2024) reported 63.58% (females) and 70.06% (males) for CAN1 alone. Similarly, Bekele *et al* (2022) noted that CAN1 explained 62% (females) and 89% (males) of the variance. The high canonical correlations observed in CAN1 (Table 8) for females (88.57%) and males (85.97%) indicate strong differentiation between chickens from different districts, which is consistent with previous findings (Melesse *et al*, 2021). However, other studies reported lower canonical correlations, where Mustefa *et al* (2021) and Muluneh *et al* (2023) reported values ranging from 50% to 80%, while Tareke *et al* (2018) and Bekele *et al* (2022) reported even lower values (30–50%).

Wilks' Lambda tests the significance of discriminant functions, with smaller values indicating greater discriminatory power (Toalombo Vargas *et al*, 2019). The lower Wilks' Lambda value of 0.16 for CAN1 in both sexes (Table 8) demonstrates strong group separation, indicating that 84% of the variability in morphometric traits arises from between-district differences rather than within-district variation. This finding aligns with other studies reporting low Wilks' Lambda values (Getachew *et al*, 2016; Melesse *et al*, 2021; Muluneh *et al*, 2023) and suggests greater between-population differentiation than studies reporting higher values (Bekele *et al*, 2022; Chebo *et al*, 2024).

The biplot based on CAN1 and CAN2 revealed distinct morphological differentiation (Figure 2). For both sexes, CAN1 was strongly loaded with size-related traits (BW, SL, BL), consistent with Melesse *et al* (2021). A clear geographic pattern emerged: East Shoa districts (particularly Lume) clustered with high CAN1 scores, reflecting larger body size, while West Arsi districts (Siraro, Shashemene) grouped with low CAN1 scores, indicating smaller dimensions. Bora district appeared distinct, supported by its high classification accuracy (Table 6). The shorter distances between districts within the same zone likely reflect shared ancestry due to non-selection, extensive gene flow resulting from continuous inbreeding, and migration over generations under traditional production systems, as noted by other authors (Ajayi *et al*, 2012; Daikwo *et al*, 2015). In contrast, the greater separation between distant groups, likely attributable to high morphometric variation and systematic sampling, offers opportunities for the conservation and genetic improvement of indigenous chickens through selective breeding strategies (Tareke *et al*, 2018).

Conclusion

This study demonstrated significant morphometric and qualitative morphological diversity among indigenous chicken populations across the studied districts. Techniques such as MCA and discriminant analysis effectively captured phenotypic distinctions, revealing clear geographic patterning. The identification of key morphometric traits, particularly shank length and body weight, as powerful discriminators and predictors makes them reliable, easy-to-measure indicators for breed characterization. Farmers and breeders can use these traits as simple, effective selection criteria for improving body conformation and market weight. These findings also underscore the need to conserve these indigenous genetic resources and utilize district-specific traits in selective breeding programmes. Future research should incorporate molecular tools to evaluate genetic admixture and validate these phenotypic observations. This work establishes a foundational basis for the sustainable use and improvement of indigenous chicken diversity in Ethiopia, contributing to enhanced food security and rural livelihoods.

Supplemental data

Supplemental Table 1. Partial contributions to inertias (eigenvalues) for the first two dimensions (Dim 1 and Dim 2)

Acknowledgments

The authors sincerely thank the farmers in the study areas for their willingness to share information and provide access to their chickens. We also extend our gratitude to the researchers who contributed to the data collection.

Author contributions

Fikrineh Negash contributed to conceptualization, methodology, data curation, formal analysis, and writing the original draft, as well as review and editing. Usman Abdulkadir contributed to methodology, project administration, resources and writing (review and editing).

Conflict of interest statement

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Ethics statement

The study was reviewed and approved by the Ethiopian Institute of Agricultural Research (EIAR) Research Review and Ethics Committee (Research code: AT/LS/Po-2015-1). Following a detailed explanation of the study's objectives and procedures, informed verbal consent was obtained from all smallholder farmers for access to their chickens.

Data availability

Data will be made available upon request.

References

Abdi, H., Valentin, D. (2007). Multiple correspondence analysis. *Encyclopedia of Measurement and Statistics* 2(4), 651–657. https://doi.org/10.4135/9781412952644.n299.

Achenef, Z. T., Ali, T. M., Mohammed, N. (2023). Phenotypic characterization of local chicken ecotypes in selected districts of North Wollo zone, Ethiopia. *East African Journal of Veterinary and Animal Sciences* 7(1), 47–56. https://doi.org/10.20372/eajvas.v7i1.473.

Ajayi, O. O., Adeleke, M. A., Sanni, M. T., Yakubu, A., Peters, S. O., Imumorin, I. G., Ozoje, M. O., Ikeobi, C. O. N., Adebambo, O. A. (2012). Application of principal component and discriminant analyses to morpho-structural indices of indigenous and exotic chickens raised under intensive management system. *Tropical Animal Health and Production* 44, 1247–1254. https://doi.org/10.1007/s11250-011-0065-1.

Aklilu, E., Kebede, K., Dessie, T., Banerjee, A. K. (2013). Phenotypic characterization of indigenous chicken population in Ethiopia. *International Journal of Interdisciplinary and Multidisciplinary Studies* 1(1), 24–32.

Alebachew, H., Abebe, A., Begna, F., Ahmed, K., Tarekegn, A., Abera, D., Mijana, D. (2019). Phenotypic characterization of local chicken ecotypes of Benishangul-Gumuz region, Western Ethiopia. *Journal of Biology, Agriculture and Healthcare* 9(7), 15–21. https://doi.org/10.7176/JBAH/9-7-03.

Assefa, H., Melesse, A. (2018). Morphological and morphometric characterization of indigenous chicken populations in Sheka Zone, South Western Ethiopia. *Poultry, Fisheries and Wildlife Sciences* 6(2), 200. https://doi.org/10.4172/2375-446X.1000200.

Balcha, Z., Baye, M., Masho, W., Admasu, Z. (2022). Morphological and morphometric features of indigenous chicken in Southwest Ethiopia. *Online Journal of Animal and Feed Research* 12(3), 132–146. https://dx.doi.org/10.51227/ojafr.2022.18.

Bayou, Y., Bayou, E., Genzebu, D., Assefa, H. (2022). Phenotypic characterization of indigenous chicken ecotypes in selected districts of Bench Maji Zone, South West Ethiopia. *International Journal of Food Science and Agriculture* 6(3), 293–300. http://dx.doi.org/10.26855/ijfsa.2022.09.009.

Begna, D., Bacha, T., Boki, S., Bekana, K. (2025). Characterization of indigenous chicken phenotypes in Liban Jawi District, Ethiopia: A qualitative and quantitative analysis. *PloS ONE* 20(1), e0307793. https://doi.org/10.1371/journal.pone.0307793.

Bekele, B., Melesse, A., Esatu, W., Dessie, T. (2022). Multivariate analysis of morphometric traits to differentiate the indigenous chicken reared under different agro-ecologies of Ethiopia. *Veterinary Integrative Sciences* 20(3): 775–791. https://doi.org/10.12982/VIS.2022.060.

Bekele, G., Goshu, G., Melesse, A., Esatu, W., Dessie, T. (2021). On-farm phenotypic and morphological characterization of indigenous chicken populations in Gambella Region, Ethiopia. *International Journal of Poultry Science* 20(1), 27–38. https://doi.org/10.3923/ijps.2021.27.38.

Belay, S., Belay, G., Resom, M. (2024). Morphometric and morphologic traits of indigenous chicken ecotypes of Tigray, northern Ethiopia. *The Scientific World Journal* 2024(1), 5537827. https://doi.org/10.1155/tswj/5537827.

Besbes, B. (2009). Genotype evaluation and breeding of poultry for performance under sub-optimal village conditions. *World's Poultry Science Journal* 65, 260–271. https://doi.org/10.1017/S0043933909000221.

Bibi, S., Noreen, S., Rehman, A., Mehmood, S., Shah, M. (2021). Morphological characteristics of native chicken of village Chhajjian, Haripur Pakistan. *Poultry Science* 100(3), 100843. https://doi.org/10.1016/j.psj.2020.11.022.

Birteeb, P. T., Boakye, T. (2020). Variant forms of qualitative traits of indigenous chickens reared under extensive system in Tolon District, Ghana. *Animal Production Science* 60(5), 705–712. https://doi.org/10.1071/AN19118.

Buranawit, K., Chailungka, C., Wongsunsri, C., Laenoi, W. (2016). Phenotypic characterization of Thai native black-bone chickens indigenous to northern Thailand. *The Thai Journal of Veterinary Medicine* 46(4), 547–554. https://doi.org/10.56808/2985-1130.2773.

Cabarles Jr, J. C., Lambio, A. L., Vega, S. A., Capitan, S. S., Mendioro, M. S. (2012). Distinct morphological features of traditional chickens (*Gallus gallus domesticus* L.) in Western Visayas, Philippines. *Animal Genetic Resources* 51, 73–87. https://doi.org/10.1017/S2078633612000410.

Camiz, S., Gomes, G. C. (2016). Alternative methods to multiple correspondence analysis in reconstructing the relevant information in a Burt's table. *Pesquisa Operacional* 36(1), 23–44. https://doi.org/10.1590/0101-7438.2016.036.01.0023.

Chebo, C., Melesse, A., Betsha, S. (2024). Quantifying phenotypic variability of indigenous chickens using morphometric traits by applying multivariate analysis: Input for sustainable rural chicken farming. *Heliyon* 10(21), e39850. https://doi.org/10.1016/j.heliyon.2024.e39850.

Chebo, C., Melesse, A., Betsha, S. (2023). Morphological descriptors of indigenous chickens: As a selection guide. *Journal of Applied Animal Research* 51(1), 729–742. https://doi.org/10.1080/09712119.2023.2274956.

Conte, G., Dimauro, C., Serra, A., Macciotta, N. P. P., Mele, M. (2018). A canonical discriminant analysis to study the association between milk fatty acids of ruminal origin and milk fat depression in dairy cows. *Journal of Dairy Science* 101(7), 6497–6510. https://doi.org/10.3168/jds.2017-13941.

Daikwo, I. S., Okpe, A. A., Ocheja, J. O. (2011). Phenotypic characterization of local chickens in Dekina. *International Journal of Poultry Science* 10(6), 444–447. https://doi.org/10.3923/ijps.2011.444.447.

Daikwo, S. I., Dike, U. A., Dim, N. I. (2015). Discriminant analysis of morphometric differences in the normal feathered and frizzle feathered chickens of north central

Nigeria. *Agro-Science: Journal of Tropical Agriculture, Food, Environment and Extension* 14(3), 12–15. http://dx.doi.org/10.4314/as.v14i3.3.

Dana, N., Dessie, T., Van der Waaij, L. H., van Arendonk, J. A. M. (2010). Morphological features of indigenous chicken populations of Ethiopia. *Animal Genetic Resources* 46, 11–23. https://doi.org/10.1017/S2078633610000652.

Dessie, T., Taye, T., Dana, N., Ayalew, W., Hanotte, O. (2011). Current state of knowledge on phenotypic characteristics of indigenous chickens in the tropics. *World's Poultry Science Journal* 67(3), 507–516. https://doi.org/10.1017/S0043933911000559.

Desta, T. T. (2021). The genetic basis and robustness of naked neck mutation in chicken. *Tropical Animal Health and Production* 53(1), 95. https://doi.org/10.1007/s11250-020-02505-1.

Desta, T. T., Dessie, T., Bettridge, J., Lynch, S. E., Melese, K., Collins, M., Christley, R. M., Wigley, P., Kaiser, P., Terfa, Z., Mwacharo, J. M. (2013). Signature of artificial selection and ecological landscape on morphological structures of Ethiopian village chickens. *Animal Genetic Resources* 52, 17–29. https://doi.org/10.1017/S2078633613000064.

Duguma, R. (2006). Phenotypic characterization of some indigenous chicken ecotypes of Ethiopia. *Livestock Research for Rural Development* 18, 131.

Egahi, J. O., Dim, N. I., Momoh, O. M., Gwaza, D. S. (2010). Variations in qualitative traits in the Nigerian local chicken. *International Journal of Poultry Science* 9(10), 978–979. https://doi.org/10.3923/ijps.2010.978.979.

Eriksson, J., Larson, G., Gunnarsson, U., Bed'Hom, B., Tixier-Boichard, M., Strömstedt, L., Wright, D., Jungerius, A., Vereijken, A., Randi, E., Jensen, P. (2008). Identification of the yellow skin gene reveals a hybrid origin of the domestic chicken. *PLoS Genetics* 4(2), e1000010. https://doi.org/10.1371/journal.pgen.1000010.

FAO (2012). Phenotypic characterization of animal genetic resources. FAO Animal Production and Health Guidelines No. 11. Rome, Italy.

FAO (2024). FAOSTAT Statistical database. https://www.fao.org/faostat/en/#data/QCL.

Fithian, W., Josse, J. (2017). Multiple correspondence analysis and the multi-logit bilinear model. *Journal of Multivariate Analysis* 157, 87–102. http://dx.doi.org/10.1016/j.jmva.2017.02.009.

Getachew, F., Abegaz, S., Assefa, A., Misganaw, M., Emshaw, Y., Hailu, A., Tessema, M., Okore, C. (2016). Multivariate analyses of morphological traits in indigenous chicken populations of Metekel zone, Northwestern Ethiopia. *Animal Genetic Resources* 59, 15–25. https://doi.org/10.1017/S2078633616000084.

Getu, A., Melaku, M., Alemayehu, Y. (2014). Phenotypic characterization of indigenous chicken ecotypes in the eastern Amahara region, Ethiopia. *Global Journal of Animal Breeding and Genetics* 2(4), 79–85.

Habimana, R., Ngeno, K., Mahoro, J., Ntawubizi, M., Shumbusho, F., Manzi, M., Hirwa, C. A., Okeno, T. O. (2021). Morphobiometrical characteristics of indigenous chicken ecotype populations in Rwanda. *Tropical Animal Health and Production* 53, 24. https://doi.org/10.1007/s11250-020-02475-4.

Hassen, H., Neser, F. W. C., van Marle-Koster, E., De Kock, A. (2007). Phenotypic variation of native chicken populations in northwest Ethiopia. *Tropical Animal Health and Production* 39, 507–513. https://doi.org/10.1007/s11250-007-9032-2.

Ige, A. O., Salako, A. E., Yakubu, A., Adeyemi, S. A. (2012). Qualitative traits characterization of Yoruba and Fulani ecotype indigenous chickens in derived savannah zone of Nigeria. *International Journal of Poultry Science* 11(10), 616–620. https://doi.org/10.3923/ijps.2012.616.620.

JMP Statistical Discovery LLC. (2022). JMP® 17 Documentation Library. Cary, NC.

Kefelegn, K. K, Feyera, B., Negassi, A. (2016). Morphometric differentiation of indigenous chicken populations of Ethiopia using discriminant analysis. In *Presentation at international research on food security, natural resource management and rural development*, Vienna, Austria.

Khangar, N. V., Kamalja, K. K. (2017). Multiple correspondence analysis and its applications. *Electronic Journal of Applied Statistical Analysis* 10(2), 432–462. http://dx.doi.org/10.1285/i20705948v10n2p432.

Lawal, R. A., Hanotte, O. (2021). Domestic chicken diversity: Origin, distribution, and adaptation. *Animal Genetics* 52(4), 385–394. https://doi.org/10.1111/age.13091.

Liswaniso, S., Mwambilwa, K., Odubote, K., Tyasi, T. L., Mweni, M., Sun, X., Xu, R., Qin, N. (2024). Zambian indigenous chicken genetic resources: phenotypic characteristics and their production systems among small-scale farmers. *Frontiers in Veterinary Science* 11, 1429869. https://doi.org/10.3389/fvets.2024.1429869.

Liyanage, R. P., Dematawewa, C. M. B., Silva, G. L. L. P. (2015). Comparative study on morphological and morphometric features of village chicken in Sri Lanka. *Tropical Agricultural Research* 26(2), 261–273. https://doi.org/10.4038/tar.v26i2.8090.

Maharani, D., Mustofa, F., Sari, A. P. Z., Fathoni, A., Sasongko, H., Hariyono, D. N. H. (2021). Phenotypic characterization and principal component analyses of indigenous chicken breeds in Indonesia. *Veterinary World* 14(6), 1665–1676. https://doi.org/10.14202/vetworld.2021.1665-1676.

Markos, S., Belay, B., Dessie, T. (2024). Morphometric differentiation of three chicken ecotypes of Ethiopia using multivariate analysis. *PLoS ONE* 19(2), e0295134. https://doi.org/10.1371/journal.pone.0295134.

McLean, D. J., Froman, D. P. (1996). Identification of a sperm cell attribute responsible for subfertility of roosters homozygous for the rose comb allele. *Biology of reproduction* 54(1), 168–172. https://doi.org/10.1095/biolreprod54.1.168.

Mekonnen, K. T., Lee, D. H., Cho, Y. G., Son, A. Y., Seo, K. S. (2023). Biometric trait characterization and multivariate discriminant analysis of traditionally bred autochthonous chickens in Ethiopia. *Agriculture* 13(11), 2142. https://doi.org/10.3390/agriculture13112142.

Melesse, A., Negesse, T. (2011). Phenotypic and morphological characterization of indigenous chicken populations in southern region of Ethiopia. *Animal Genetic Resources* 49, 19–31. https://doi.org/10.1017/S2078633611000099.

Melesse, A., Tadele, A., Assefa, H., Taye, K., Kebede, T., Taye, M., Betsha, S. (2021). Assessing the morphological diversity of Ethiopian indigenous chickens using multivariate discriminant analysis of morphometric traits for sustainable utilization and conservation. *Poultry Science Journal* 9(1): 61–72. https://doi.org/10.22069/psj.2021.18469.1644.

Moreda, E., Singh, H., Sisaye, T., Johansson, A. M. (2014). Phenotypic characterization of indigenous chicken population in South West and South part of Ethiopia. *British Journal of Poultry Sciences* 3(1), 15–19. https://doi.org/10.5829/idosi.bjps.2014.3.1.8370.

Moto, E., Rubanza, C. D. (2023). Phenotypic characterization of indigenous chicken in the central zone of Tanzania. *Open Agriculture* 8(1), 20220218. https://doi.org/10.1515/

opag-2022-0218.
Moula, N., Farnir, F., Salhi, A., Iguer-Ouada, M., Leroy, P., Antoine-Moussiaux, N. (2012). Backyard poultry in Kabylie (Algeria): from an indigenous chicken to a local poultry breed? *Animal Genetic Resources* 50, 87–96. https://doi.org/10.1017/S207863361200001X.
Muluneh, B., Taye, M., Dessie, T., Wondim, D. S., Kebede, D., Tenagne, A. (2023). Morpho-biometric characterization of indigenous chicken ecotypes in north-western Ethiopia. *PLoS ONE* 18(6), e0286299. https://doi.org/10.1371/journal.pone.0286299.
Mustefa, A., Kenfo, H., Belayhun, T., Hailu, A., Assefa, A. (2021). Morphometric and morphological characterization of chicken resources adapted to pastoral and agropastoral areas of southern Ethiopia. *Genetic Resources* 2(4), 72–84. https://doi.org/10.46265/genresj.NDFM2712.
Navara, K. J., Anderson, E. M., Edwards, M. L. (2012). Comb size and color relate to sperm quality: A test of the phenotype-linked fertility hypothesis. *Behavioral Ecology* 23(5), 1036–1041. https://doi.org/10.1093/beheco/ars068.
Negassa, D., Melesse, A., Banerjee, S. (2014). Phenotypic characterization of indigenous chicken populations in Southeastern Oromia Regional State of Ethiopia. *Animal Genetic Resources* 55, 101–113. https://doi.org/10.1017/S2078633614000319.
Ngeno, K., Vander Waaij, E. H., Kahi, A. K. (2014). Indigenous chicken genetic resources in Kenya: Their unique attributes and conservation options for improved use. World's Poultry Science Journal 70, 173–184. https://doi.org/10.1017/S0043933914000154.
Nigussie, H., Kebede, K., Ameha, N. 2015. Phenotypic and morphological characterization of indigenous chicken populations in southern zone of Tigray, Ethiopia. *Journal of Biology, Agriculture and Healthcare* 5(21), 132–141.
Otecko, N. O., Ogali, I., Mauki, D. H., Ogada, S., Moraa, G. K., Lichoti, J., Agwanda, B., Peng, M. S., Ommeh, S. C., Zhang, Y. P. (2019). Phenotypic and morphometric differentiation of indigenous chickens from Kenya and other tropical countries augments perspectives for genetic resource improvement and conservation. *Poultry Science* 98(7), 2747–2755. http://dx.doi.org/10.3382/ps/pez097.
Owens, I. P., Hartley, I. R. (1998). Sexual dimorphism in birds: Why are there so many different forms of dimorphism? *Proceedings of the Royal Society B: Biological Sciences* 265(1394), 397–407. https://doi.org/10.1098/rspb.1998.0308.
Picardal, J. P., Afable, F. A., Lagman, M. C. A., Campoto, E. A., Palada, E. P., Valdez Jr, M. B. (2015). Phenotypic characterization of native chickens (*Gallus gallus domesticus*) in Eastern Samar, Philippines. *IAMURE International Journal of Ecology and Conservation* 15, 242–265. http://dx.doi.org/10.7718/ijec.v15i1.1005.
Price-Waldman, R., Stoddard, M. C. (2021). Avian coloration genetics: Recent advances and emerging questions. *Journal of Heredity* 112(5), 395–416. https://doi.org/10.1093/jhered/esab015.
Rotimi, E. A., Egahi, J. O., Adeoye, A. A. (2016). Phenotypic characterization of indigenous chicken population in Gwer-West, Benue State, Nigeria. *World Scientific News* 3(53), 343–353.
Sarker, N. R., Hoque, A., Faruque, S., Islam, N., Bhuiyan, F. H. (2014). An ex situ study on body characteristics and effect of plumage color on body weight of indigenous chicken (*Gallus domesticus*) in Bangladesh. *Acta Scientiarum. Animal Sciences* 36, 79–84. https://doi.org/10.4025/actascianimsci.v36i1.20118.
SAS Institute Inc. (2022). JMP® Pro Version 17.0.0. Cary, NC.
Semakula, J., Lusembo, P., Kugonza, D. R., Mutetikka, D., Ssennyonjo, J., Mwesigwa, M. (2011). Estimation of live body weight using zoometrical measurements for improved marketing of indigenous chicken in the Lake Victoria basin of Uganda. *Livestock Research for Rural Development* 23(8), 170.
Sourial, N., Wolfson, C., Zhu, B., Quail, J., Fletcher, J., Karunananthan, S., Bandeen-Roche, K., Béland, F., Bergman, H. (2010). Correspondence analysis is a useful tool to uncover the relationships among categorical variables. *Journal of Clinical Epidemiology* 63(6), 638–646. https://doi.org/10.1016/j.jclinepi.2009.08.008.
Suyatno, S., Sujono, S., Winaya, A., Zalizar, L., Pangestu, M. (2023). Characterization of qualitative and quantitative traits of four types of Indonesian native chickens as ancestor of new strains of local super laying hens. *Jordan Journal of Biological Sciences* 16(2), 71–179. https://doi.org/10.54319/jjbs/160201.
Tadese, D., Bekele, B., Hailemesikel, D., Wolde, B., Esatu, W., Dessie, T. (2024). On-farm phenotypic characterization of indigenous chicken populations and their productive and reproductive performances in central Ethiopia. *Journal of Livestock Science* 15(15), 181–189. https://doi.org/10.33259/JLivestSci.2024.181-189.
Tamirat, D., Getachew, T., Masho, W., Admasu, Z. (2023). Morphological and morphometric features of indigenous chicken in North Shewa zone, Oromia regional state, Ethiopia. *The Scientific Temper* 14(3), 609–617. https://doi.org/10.58414/SCIENTIFICTEMPER.2023.14.30.07.
Tareke, M., Assefa, B., Abate, T., Tekletsadik, E. (2018). Evaluation of morphometric differences among indigenous chicken populations in Bale zone, Oromia Regional State, Ethiopia. *Poultry Science Journal* 6(2), 181–190. https://doi.org/10.22069/psj.2018.14974.1329.
Terfa, Z. G., Garikipati, S., Kassie, G. T., Dessie, T., Christley, R. M. (2019). Understanding farmers' preference for traits of chickens in rural Ethiopia. *Agricultural Economics* 50(4), 451–463. https://doi.org/10.1111/agec.12502.
Tixier-Boichard, M., Ayalew, W., Jianlin, H. (2008). Inventory, characterization and monitoring. *Animal Genetic Resources Information* 42, 29–44. https://doi.org/10.1017/S1014233900002534.
Toalombo Vargas, P. A., Navas Gonzalez, F. J., Landi, V., Leon Jurado, J. M., Delgado Bermejo, J. V. (2019). Sexual dimorphism and breed characterization of Creole hens through biometric canonical discriminant analysis across Ecuadorian agroecological areas. *Animals* 10(1), 32. https://doi.org/10.3390/ani10010032.
Traoré, G. F., Tapsoba, A. S. R., Bayala, B., Soudré, A., Traore, A., Dayo, G. K., Ávarez, I., Fernández, I., Sanou, M., Tamboura, H. H., Goyache, F. (2018). Multivariate analyses of morphological traits characterize the Guinea fowl (Numida meleagris) of Burkina Faso as a homogeneous population. *European Poultry Science* 82, 1–13. https://doi.org/10.1399/eps.2018.251.
Udeh, I., Ugwu, S. O. C., Ogagifo, N. L. (2011). Predicting semen traits of local and exotic cocks using linear body measurements. *Asian Journal of Animal Sciences* 5(4), 268–276. https://doi.org/10.3923/ajas.2011.268.276.
Veflen, N., Storstad, O., Samuelsen, B., Langsrud, S., Hagtvedt, T., Ueland, Ø., Gregersen, F., Scholderer, J. (2017). Food scares: Reflections and reactions. *International Journal on Food System Dynamics* 8(2), 155–164. http://dx.doi.

org/10.18461/ijfsd.v8i2.826.

Wario, D. D., Tadesse, Y., Yadav, S. B. S. (2021). On-Farm phenotypic characterization of indigenous chicken, in Dire and Yabello districts, Borena zone, Oromia Regional State, Ethiopia. *Journal of Genetic Resources* 7(1), 36–48. https://doi.org/10.22080/jgr.2020.19954.1211.

Yaemkong, S., Phromnoi, S., Mingchai, C., Jongjitvimol, T. (2024). Characterization of phenotypic variation in indigenous chicken populations in lower Northern Thailand to improve chicken breeding. *International Journal of Zoology* 2024(1), 9985076. https://doi.org/10.1155/2024/9985076.

Yakubu, A., Ari, M. M. (2018). Principal component and discriminant analyses of body weight and conformation traits of Sasso, Kuroiler and indigenous Fulani chickens in Nigeria. *The Journal of Animal and Plant Sciences* 28(1), 46–55.

Yussif, I., Kugonza, D. R., Okot, M. W., Amuge, P. O., Costa, R., Dos Anjos, F. (2023). Uganda chicken genetic resources: I. phenotypic and production characteristics. *Frontiers in Genetics* 13, 1033031. https://doi.org/10.3389/fgene.2022.1033031.

Genetic Resources

ORIGINAL ARTICLE

Genetic Resources (2026), 7 (13), 29–40
DOI: 10.46265/genresj.SARW4599
https://www.genresj.org
ISSN: 2708-3764

Phenotypic diversity of indigenous goats across three agroecological zones in southeastern Ethiopia

Kebede Tilahun[a,b,*], Aberra Melesse[a], Simret Betsha[a]

[a] *College of Agriculture, School of Animal and Range Sciences, Hawassa University, P.O. Box 05, Hawassa, Ethiopia*
[b] *College of Agriculture, Department of Animal Science, Kebridehar University, P.O. Box 250, Kebridehar, Ethiopia*

Abstract: The study aimed to evaluate the phenotypic diversity of indigenous goats reared in three agroecological zones of southeastern Ethiopia based on their morphometric traits. Multi-stage purposive and random sampling methods were applied to select the study areas and the goats. A total of 601 (463 female and 138 male) goats were randomly sampled, measured and described using 16 morphometric traits. Data collected were analyzed using univariate and multivariate statistical procedures in SAS version 9.4. Univariate analysis revealed significant variations of all morphometric traits across the agroecological zones. The goats in the lowland agroecological zone were larger ($p < 0.05$) than those reared in other agroecological zones. Goats reared in the highland agroecological zone were smaller and lighter in size, while those from the midland agroecological zone showed intermediate body size and weight. Multivariate analysis (i.e. canonical discriminant and cluster analysis) showed that the goats in the three agroecological zones were different in terms of their morphology, with the largest Mahalanobis distance (42.3%) being observed between lowland and highland goat populations. Discriminant analysis correctly assigned 86% of goats to their source populations. These results indicate the existence of morphological diversity among goat populations in the three agroecological zones of southeastern Ethiopia and suggest the need to develop conservation and breeding strategies aimed at retaining the observed diversity at the phenotypic level. Further studies using molecular tools are needed to elucidate the observed diversity at the phenotypic level and to design appropriate strategies for the sustainable management of these animal genetic resources.

Keywords: Agroecological zone, conservation, diversity, goat, morphometric trait, multivariate analysis

Citation: Tilahun, K., Melesse, A. and Betsha, S. (2026) "Phenotypic diversity of indigenous goats across three agroecological zones in southeastern Ethiopia", Genetic Resources, 7(13), pp. 29–40. doi: 10.46265/genresj.SARW4599.

Introduction

Genetic diversity of livestock species across various environments is vital for their long-term viability and resilience in the face of climate change (Kuthu *et al*, 2022) and helps animals to adapt to changing environments and demands (Razgour *et al*, 2019; Velado-Alonso *et al*, 2020; Shuma *et al*, 2024). It provides a baseline for survival, productivity and genetic improvement of animal population in diverse climates (Oladepo *et al*, 2017; Ganiyu *et al*, 2018). Genetic diversity in indigenous populations is essential for the sustainable utilization and conservation of resources (Velado-Alonso *et al*, 2020; Habimana *et al*, 2021; Sheriff *et al*, 2021). Morphological variation among indigenous goats has important economic and socio-cultural value (Getaneh *et al*, 2022; Melesse *et al*, 2022) and is a good indicator of the selection regime and history of the breed (Mdladla *et al*, 2017). Indigenous breeds are an invaluable and irreplaceable source of genetic diversity (Maksimovic *et al*, 2023). Proper characteriztion at phenotypic and genetic levels is essential for their improvement and conservation (Whannou *et al*, 2021; Worku and Melesse, 2021; Okoro *et al*, 2023).

Characterization involves three types of information, namely phenotypic, genetic and historical (FAO, 2012). Phenotypic characterization provides essential information for proper management of farm animal genetic resources at local, national, regional and global levels (FAO, 2012)

* Corresponding author: Kebede Tilahun (keedetilahun2000@gmail.com)

Received: 07.02.2025 | Accepted: 20.11.2025 | Published online: 27.01.2026

and is important to describe the uniqueness of animal genetic resources (Sako *et al*, 2024). Morphological characterization includes both the description of qualitative and quantitative characteristics (Deribe *et al*, 2021; Kitila *et al*, 2025), and it is the first step in identifying unique traits (Kandoussi *et al*, 2021; Melesse *et al*, 2022) and population structure, especially in resource-poor areas, where advanced genomic tools are limited (Bousbia *et al*, 2021; Nguluma *et al*, 2022; Akounda *et al*, 2023). Morphometric traits specifically serve as reliable indicators for breed identification and the productivity potential of livestock resources (Aliyu *et al*, 2021; Elzarei *et al*, 2023; Selvan *et al*, 2023). They determine the growth, development, shape and proportion of animals, and have a strong correlation with production traits (Ghahri *et al*, 2019; Chokoe *et al*, 2020; Muhammad *et al*, 2021; Kuthu *et al*, 2022; Kitila *et al*, 2025). For example, traits such as body weight, body height and chest girth are direct indicators of body size and are important growth determinants that affect economic viability (Valsalan *et al*, 2020). These traits have moderate to high heritability, making them suitable for selection in genetic improvement programmes (Sarma *et al*, 2019) and have diverse implications in designing genetic improvement programs (Melesse *et al*, 2022; Akounda *et al*, 2023).

Ethiopia is endowed with large and diversified goat breeds/populations (Shuma *et al*, 2024), which are raised in different production systems and agroecological zones. These goat populations exhibit significant phenotypic and genetic diversity resulting from environmental pressure and traditional breeding practices (Tarekegn *et al*, 2021). The existence of such large goat populations, coupled with their diverse agroecological conditions, production and husbandry systems, justifies the need for characterization. A lack of characterization can lead to improper utilization, indiscriminate crossbreeding and dilution, and replacement of a local population without knowing their unique genetic merit. These phenomena result in a loss of genetic diversity and population uniformity, which has become a global concern (Arsoy *et al*, 2024). This is especially true in developing nations, where indigenous breeds are replaced at an alarming rate by high-producing exotic breeds irrespective of their genetic potential. Thus, the findings of characterization studies would enable breeders and policymakers to take appropriate actions to safeguard local animal genetic resources from genetic erosion through adulteration and uncontrolled crossbreeding activities.

The goat population distributed in the southeastern part of Ethiopia is commonly referred to as the Arsi-Bale breed/type. They are reared in the highlands of Arsi and Bale and in the lowlands of the Sidama region. Different attempts have been made to characterize this goat population. For example, Hankamo *et al* (2020) found the existence of remarkable differences among the goats in the Sidama region. Muluneh and Tadesse (2022), in their study, detected variation between goats in two districts of southern Ethiopia. Abebe and Korato (2020) detected phenotypic variation among the goats in the Arsi zone. In addition, Guyo *et al* (2023) found the divergence of goats reared in three agroecological zones of the Bale zone. The inconsistency and divergence of the findings of these authors suggest that a goat population is underrepresented. In addition, none of these authors have considered agroecological zones and multivariate techniques to analyze the morphometrical data except Guyo *et al* (2023). Furthermore, there is considerable population movement between the Arsi-Bale and nearby goat populations, which could lead to phenotypic differences and genetic adulterations of goats reared in these areas (Worku and Melesse, 2021). As a result, characterizing and evaluating genetic diversity and population structure is imperative to design suitable breeding programmes, allowing sustained genetic improvement and conservation of the goat populations in their native environments. Therefore, this study aimed to characterize the phenotypic diversity of local goat populations reared in three agroecological zones to generate baseline data for designing sustainable genetic improvement strategies of local goat populations in the study areas.

Materials and methods

Study area

This study was conducted in three agroecological zones of southeastern Ethiopia, namely lowland, midland and highland. These agroecological zones were identified in two major regions, the Sidama and Oromia regions. The midland and lowland agroecological zones were sampled from the Sidama region, while the highland zone was considered from the West Arsi zone of the Oromia region (Figure 1). The description of agroecological zones is summarized in Table 1.

Sampling procedure

Secondary data on goat distribution and climate were obtained from zonal livestock and fishery development offices. These data were used for the selection of the three agroecological zones considered in this study according to their gradient and the abundance of goats. Multi-stage purposive sampling methods were employed to select districts, kebeles (i.e. the smallest administrative unit in Ethiopia), and households. Loka Abaya, Aleta Chucko, and Kofele districts were specifically selected to represent the lowland, midland and highland agroecological zones, respectively. In these nine rural kebeles, a total of 252 households (76 from lowland, 84 from midland and 92 from highland, agroecological zones) were selected based on the size of their goat herds. From each of these households, two to three goats were sampled, and morphometric measurements were taken. The sample size was then estimated using Yamane's (1967) formula described as follows:

$$\mathbf{n} = \frac{\mathrm{N}}{1+\mathrm{N}(\mathrm{e})^2}$$

where n is the sample size, N is the population size (108,600 in the lowland, 36,439 in the midland, and 3,718 in the highland agroecological zones), and e is the level of precision (7%). Accordingly, a total of 601 adult goats (i.e. 204 from lowland, 203 from midland and 194 from highland agroecological zones) with at least one pair of permanent incisors were randomly sampled.

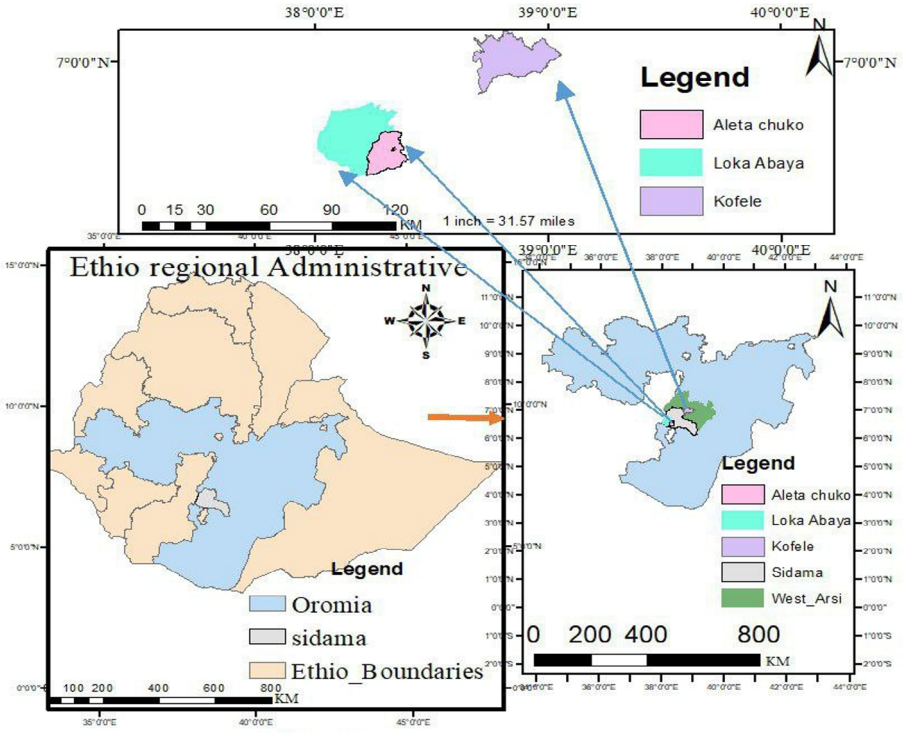


Figure 1. Map of the study area

Table 1. Description of the agroecological zones with coordinates and climates. Source: Hankamo *et al*, 2020; Worku and Melesse 2021; Chebo *et al*, 2023.

Agroecological zones	District	Latitude	Longitude	Average temperature (°C)	Rainfall (mm)	Altitude (m a.s.l.)
Highland	Kofele	06°50'–07°09' N	38°38'–39°04' E	5–17	1,500–2,100	2,000–3,050
Midland	Aleta Chuko	6°46'–7°10' N	38°04'–38°24' E	18–28	1,200–1,400	1,400–2,300
Lowland	Loka Abaya	6°42'–7°83' N	37°92– 39°14' E	10–32	900–1,400	1,001–2,000

Morphometric data collection

Morphometric data were collected between October 2023 and February 2024 in the three agroecological zones. Sixteen morphometric traits were recorded from 601 goats of both sexes (Supplemental Table 1). Measurements were as follows: body weight (BW), taken using a suspended balance; chest girth (CG), measured as a circumference of the chest just behind the forelegs; rump length (RL), as distance from hip (Tuber coxa) to the pin (Tuber ischi); rump width (RW), as horizontal distance between the extreme lateral points of the hook bone of pelvis; rump height (RH), measured as a distance from the ground to the highest point of rump; height at withers (HW), as a distance from the ground to the withers; body length (BL), measured as the distance from the point of shoulder to the pin bone; chest width (CW), as the width of the chest between the briskets; paunch girth (PG), as the circumference of the belly at the centre; ear length (EL), measured from the base to the tip of the ear; ear width (EW), measured from one tip of the ear to the other across the centre of the ear; head length (HdL), the length of front head from middle of top head between horn site to end of mouth; head width (HdW), as the width of front head from the base of left and right ear; chest depth (CD), the vertical distance from sternum (bottom brisket surface) to withers; shoulder width (SW), as the distance between the tip of two withers; and fore cannon circumference (FCC), measured as a circumference at the narrowest part of the bone jointing fetlock and knee joint. The body weight of the goat was measured using a 50kg portable digital scale, while linear body measurement traits were measured using plastic tape graduated in cm. All measurements were taken early in the morning, before the goats were released for feeding. Measurements were taken while the animals were standing on a flat surface and held by one person to reduce measurement error. Pregnant females were excluded from the measurements.

Statistical analysis

Univariate analysis

The morphometric data were analyzed using the General Linear Model Procedures of the Statistical Analysis System (PROC GLM of SAS version 9.4, 2016) to identify the effect of agroecological zone on morphometric traits. When the F-test declared significant, multiple least square means were compared using the Tukey-Kramer test. The univariate model used to analyze the effect of an agroecological zone is as follows:

$Y_i = \mu + A_i + e_i$

Y_i = response of the observed dependent variables

μ= the overall mean

A_i = the fixed effect of i^{th} agroecological zones (i = highland, midland and lowland)

e_i = residual error

Multivariate analysis

The stepwise regression procedure was conducted on data collected from female and male goats separately to identify the best-fitted regression model for the prediction of live body weight of male and female goats separately. The identified model was then plotted to generate a graph using the procedure of linear regression. The stepwise multiple linear regression was carried out to predict the body weight from linear traits using the model:

$$Y_j = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 \ldots \beta_{16} X_{16} + e_j$$

Where:

Y_j = the response variable (BW)

β_0 = the intercept

$X_1, X_2 \ldots X_{16}$ = linear traits

$\beta_1, \beta_2 \ldots \beta_{16}$ = regression coefficients of the $X_1, X_2 \ldots X_{16}$

e_j = residual error

Sixteen morphometric traits were subjected to the stepwise discriminant analysis (PROC STEPDISC) to identify the morphometric traits that have high discriminating power. However, the analyses have determined that all morphometric traits have discriminating power. The procedure of cluster analysis (PROC CLUSTER) was applied to construct a dendrogram using the average linkage method to cluster the goat populations into their morphometric similarity. Moreover, the canonical discriminant analysis (PROC CANDISC) was applied to determine the Mahalanobis distance, univariate and multivariate statistics, and canonical variables. The TEMPLATE and SGRENDER procedures were used to create a plot of the first two canonical variables in a scatter graph for visual interpretation. Furthermore, the quadratic discriminant analysis (PROC DISCRIM) was conducted to determine the probabilities of classifying individual animals in their actual source population.

Results

Univariate analysis

Agroecological zone had a considerable effect ($p < 0.05$) on all morphometric traits except PG and CW (Table 2). The lowland goats showed significantly higher BW and height traits ($p < 0.001$) than those of their midland and highland counterparts. The midland goats had higher HdL than those reared in the other two agroecological zones. The goats in the highland agroecology had a comparatively smaller height and EL than those of the lowland and midland agroecological zones (Table 2). Relatively lower coefficients of variation (4.94–15.4%) were observed in goats reared in the highland zone, while larger CV values were observed for those of the midland (7.41–21.8%) and lowland (7.20–22.2%) agroecological zones.

Table 2. Least squares means (± SE) and coefficients of variation for body weight (kg) and linear morphometric traits (cm) of goats reared across three agroecological zones. Least square means within a row with different superscripts are significantly different ($p < 0.05$); CV, coefficients of variations; N, number of goats; BW, body weight; CG, chest girth; HW, height at withers; BL, body length; RH, rump length; PG, paunch girth; CW, chest width; CD, chest depth; SW, shoulder width; RL, rump length; RH, rump height; HdL, head length; HdW, head width; EL, ear length; EW, ear width and FCC, fore canon circumference.

Traits	Highland (N = 194)	CV	Midland (N = 203)	CV	Lowland (N = 204)	CV	P values
BW	25.6b±0.45	15.4	27.2b± 0.31	21.8	28.3a±0.32	22.2	<.001
CG	71.6a±0.49	6.94	70.8b±.0.34	8.59	72.1a± 0.35	9.64	0.020
HW	63.9b ±0.40	5.28	66.6a±0.27	7.49	67.3a± 0.29	7.78	<.001
BL	63.4b±0.45	7.20	64.2b ± 0.31	8.21	65.5a±0.32	8.82	0.001
RH	64.7b±0.37	4.94	66.7a ± 0.25	6.93	66.9a±0. 26	6.61	<.001
PG	80.0±0.74	8.97	81.2± 0.51	11.6	79.9±0.53	10.3	0.133
CW	18.5± 0.21	11.2	18.2± 0.21	12.7	18.1±0.15	13.1	0.190
CD	30.9a± 0.22	7.77	31.0a± 0.15	7.62	30.3b± 0.16	9.91	0.003
SW	18.8a ± 0.17	7.89	18.9a± 0.12	10.5	17.8b± 0.12	12.5	<.001
RL	20.0 b± 0.15	7.72	19.9b± 0.10	9.49	20.5a±0.11	8.62	<.0001
RH	13.9b± 0.12	10.4	13.7b± 0.08	9.85	14.4a±0.09	11.1	<.001
HdL	20.5b± 0.14	10.1	21.3a ± 0.10	7.54	20.9b± 0.10	7.20	<.001
HdW	15.0a ± 0.15	12.1	14.5b ± 0.10	11.4	14.4b± 0.11	14.3	0.001
EL	12.7c± 0.13	6.75	13.5b± 0.09	9.26	14.3a± 0.09	9.26	<.001
EW	7.4a± 0.06	6.95	7.5a ± 0.04	7.41	7.2b±0.05	8.30	<.001
FCC	7.6b±0.06	7.11	7.7b± 0.04	10.1	7.9a±0.04	10.3	<.001

Multivariate analyses

Prediction of live body weight

The correlation analysis has shown a strong association between CG and BW, indicating that the live weight of female and male goats could be best estimated from their CG. The equation to estimate live body weight of female and male goats is indicated in Figures 2A and B, respectively. Likewise, CG could be also used to estimate live weight in combined female and male data (Supplemental Figure 1).

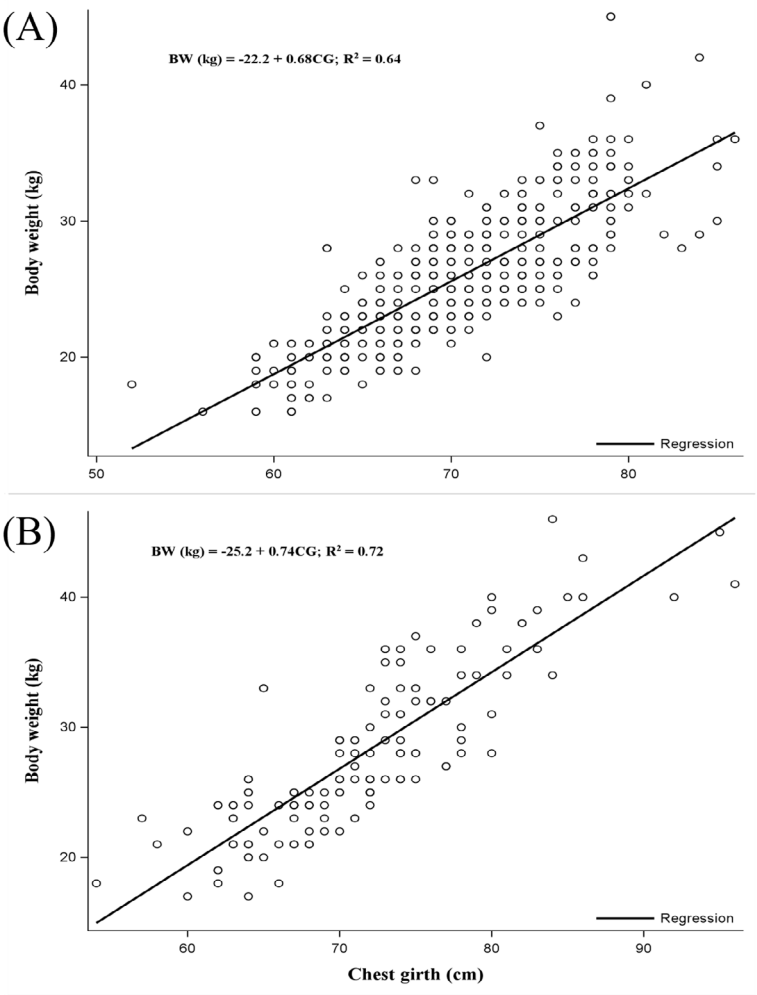


Figure 2. Linear regression plots showing the prediction equation of live body weight of female (A) and male goats (B) using morphometric traits. BW, body weight (kg); CG, chest girth (cm).

Canonical discriminant analysis

All the squared Mahalanobis distances based on morphometric traits were significant ($p < 0.001$), and the longest (42.3) were observed between highland and lowland agroecological zones (Table 3). Univariate analysis (ANOVA) revealed that all morphometric traits significantly ($p < 0.05$) contributed to the total variation, except body weight and body length (Supplemental Table 2).

As shown in Table 4, all multivariate statistics were highly significant ($p < 0.001$) between the three agroecological zones. The Wilk's lambda showed 69% of the variability was due to variations between populations rather than within a population (Table 4).

Table 3. Mahalanobis distance between goat populations reared in the three agroecological zones

Agroecological zones	Midland	Highland	Lowland
Midland	0	17.6	26.1
Highland	-	0	42.3
Lowland	-	-	0

Table 4. Multivariate statistics and F approximation of goat populations in the three agroecological zones

Statistic	Value	F Value	Num DF	Den DF	Pr > F
Wilks' Lambda	0.31	26.9	34	1,156	<.0001
Pillai's Trace	0.85	25.3	34	1,158	<.0001
Hotelling-Lawley Trace	1.69	28.6	34	1,035.8	<.0001
Roy's Greatest Root	1.27	43.3	17	579	<.0001

Both CAN1 and CAN2 explained 75% and 25% of the total variance, respectively (Table 5). The highest (74%) canonical correlation was observed for CAN1. The total standardized canonical coefficients indicate that CD and SW significantly contributed to CAN1, while HdL and PG contributed to CAN2 (Supplemental Table 3). Shoulder width and CD were the most important discriminating traits in CAN1, while HdL and RH were for CAN2 (Supplemental Table 4).

Figure 3 showed that CAN1 was the best at discriminating between highland and lowland goat populations, while CAN2 slightly differentiate the midland goats from the other two populations but not strongly. The distribution of the goat populations in the three agroecological zones was clear and apparent. Accordingly, the highland goat populations are distributed to the right side of the graph to CAN1 while those in the lowland were clustered to the left side of the graph to CAN1. The midland goat population occupied the centre and they overlap with both highland and lowland agroecological zones.

Table 5. Summary of canonical correlation, eigenvalue and likelihood ratio. CAN1, canonical variable 1; CAN2, canonical variable 2.

Function	Canonical correlation	Eigenvalues			Likelihood ratio	Approximate F-value	Pr > F
		Eigenvalue	Proportion	Cumulative			
CAN 1	0.74	1.27	0.75	0.75	0.31	26.9	<.0001
CAN 2	0.53	0.41	0.25	1.00	0.71	15.0	<.0001

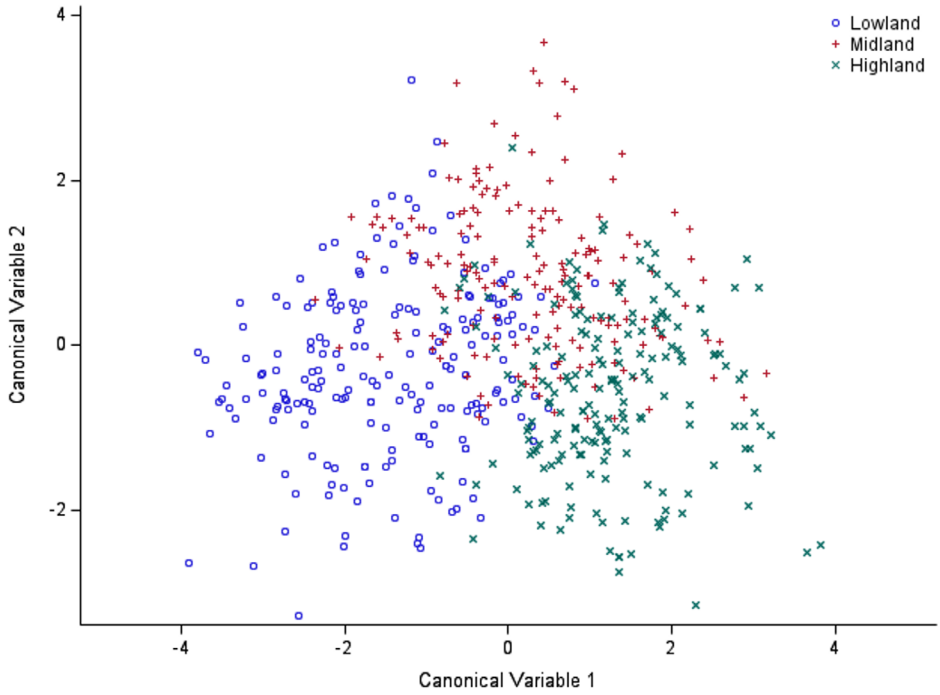


Figure 3. Canonical discriminant plot showing group discrimination among agroecological zones

Cluster analysis

As presented in Figure 4, the dendrogram clearly clustered the goat population of the lowland agroecological zone separately as a distinct group, while those reared in the highland and midland agroecological zones were closely clustered as subclusters.

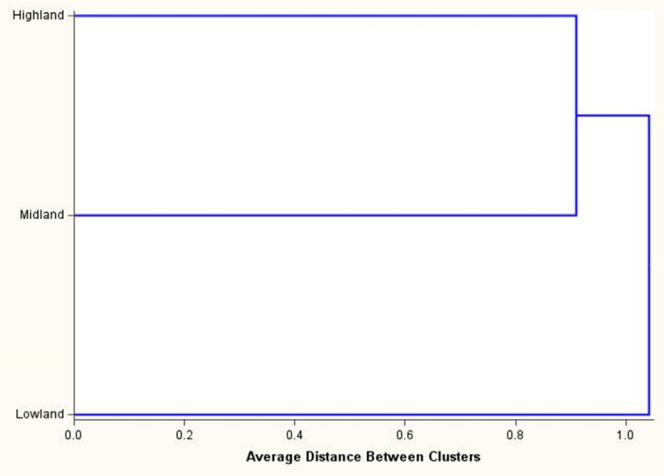


Figure 4. Dendrogram showing classification of the goat population in the three agroecological zones

Discriminant analysis

The quadratic discriminant analysis indicated that 85.6% of the goat populations in the midland agroecological zone were accurately classified into their source population while the rest (14.4%) were misclassified into highland and lowland (Table 6). About 88% of the lowland populations were correctly assigned to their source population, while about 12% of them were misclassified to the other agroecological zones. The accuracy of classification was tested with re-substitution and cross-validation options, and the classification was successful with an overall rate of 78% (Table 6).

Discussion

Univariate analysis

Assessing the phenotypic diversity of indigenous goat populations is crucial for genetic improvement and conservation purposes. The phenotypic difference among the goat populations is due to variation in agroecological zone as a result of variability in climate and vegetation. The variation in the body size of animals could be affected by management and environmental factors (Nantongo *et al*, 2024), nutritional variation (Monau *et al*, 2018; Getahun *et al*, 2020), availability of feed and its type (Ofori *et al*, 2021), variation in rainfall pattern, vegetation and diurnal temperature (Ntonga *et al*, 2025), and low artificial selection intensity (Melesse *et al*, 2022). Morphometric traits are associated with productivity and provide important information on the suitability of animals for selection (Bousbia *et al*, 2021; Tenagne *et al*, 2023). The phenotypic variability within indigenous goats indicates ecological adaptation and is important for selection and breeding programme design tailored to specific agroecological zones and production systems. The observed high height in goats reared in the lowland and midland agroecological zones is in accordance with the results of Guyo *et al* (2023), who reported similar findings. The larger FCC in the lowland goat population is an indication of good skeletal development to carry their body mass while browsing in mountainous terrain and is consistent with the reports of Dea *et al* (2019) for the lowland agroecological zone of the Gamo-Gofa zone. The observed FCC in this study is also in line with that of Maksimović *et al* (2023), who reported comparable results for Serbian White and Balkan goat breeds. The long ear of the lowland goat indicates its adaptation to the hot lowland areas, in which it helps to dissipate excess heat during hot periods.

On the other hand, the smallest HW and RH observed in the highland goat population indicate that their smaller body mass is close to the ground, which might be beneficial to adapt to the cold environment. The lowest body measurements observed in the highland goat population

Table 6. Percentages and numbers of individuals classified according to their source of origin based on morphometric traits.

Agroecological zones	Midland	Highland	Lowland	Total
Midland	85.6% (173)	5.9% (12)	8.4% (17)	100% (202)
Highland	12.5% (24)	84.9% (163)	2.6% (5)	100% (192)
Lowland	9.9% (20)	2.5% (5)	87.6% (178)	100% (203)
Error count rate	0.14	0.15	0.12	0.14
Cross validation				
Midland	76.2% (154)	9.9% (20)	13.9% (28)	100% (202)
Highland	18.8% (36)	78.1% (150)	3.1% (6)	100% (192)
Lowland	16.8% (34)	4.9% (10)	78.3% (159)	100% (203)
Error count rate	0.24	0.22	0.22	0.22

might be attributed to a feeding regime affected by cold stress, the shrinkage of grazing area (Guyo *et al*, 2023) and the absence of browsing pasture. The smaller RL and RW in midland goats may indicate their slimmer body and lower meat production potential as compared with the two others. The observed difference between the three agroecological zones may result from both environmental influences and underlying genetic variations due to long-term adaptation to the respective production environments. The variability in morphometric traits among agroecological zones indicates the adaptive potential of indigenous goats and suggests their ability to survive in diverse environments.

On the other hand, based on their HW, the goat populations in the three agroecological zones could be classified into two groups as larger and smaller breed types. Goats having HW over 65cm are considered a larger breed type, while those between 51 and 65cm are classified as smaller breed types (Laouadi *et al*, 2020). In the current study, the HW of 66.6cm and 67.3cm observed in the midland and lowland goat populations indicates that they are larger breed types. On the other hand, goats reared in the highland agroecological zone had comparatively lower HW (63.9cm) and could be classified as smaller breed types (Laouadi *et al*, 2020). The CV for different traits showed relatively low variability, indicating that the goat population within each agroecological zone was characterized by possessing uniform morphometrical traits. The CV observed in the highland goats is comparable with those reported for Hararghe highland goats (Takele *et al*, 2021), Arsi-Bale goats (Guyo *et al*, 2023), central Tigray goats (Birhaniea *et al*, 2019), and East Gojjam goats (Getahun *et al*, 2020). In contrast, lower CV values were reported for indigenous goats reared in five administrative zones of Ethiopia (Melesse *et al*, 2022).

Multivariate analysis

Multivariate analysis of morphometric traits is effective in determining variations at phenotypic and genetic levels (Rotimi *et al*, 2020; Melesse *et al*, 2022) and is suitable for exploring breed structure and diversity (Selvan *et al*, 2023). Among morphometric traits, BW is an important economic trait in determining growth performance, market price and management aspects, including drug dosage in different classes of age groups. However, measuring the live weight of animals in rural areas of Africa is quite challenging (Vanvanhossou *et al*, 2018). This difficulty arises mainly due to the lack of adequate measuring scales under field conditions, highlighting the need for alternative methods of estimating animal weight. Purchasing weighing scales for BW measurement is much more expensive than obtaining a textile meter tape for measuring linear morphometric traits. Therefore, linear measurements can serve as indirect indicators of live weight and carcass traits (Mebratie *et al*, 2022). They also play a vital role in effective management and selection of replacement stock (Takele *et al*, 2021; Getaneh *et al*, 2022; Kuthu *et al*, 2022). For example, studies have shown that CG and BL are highly correlated with live BW in various goat breeds (Mebratie *et al*, 2022). Therefore, in areas where weighing equipment is not affordable, predicting the BW of goats from linear traits is not only practical but also essential for improving livestock management and supporting rural livelihoods.

In the current study, the CG was identified as a suitable variable to predict the BW of goats, in line with the reports of Hankamo *et al* (2020), Tade *et al* (2021), Takele *et al* (2021), and Tyasi and Tada (2023). Dea *et al* (2019) reported the estimation of BW from the combinations of HW, BL and CG. However, the inclusion of many traits to estimate the live BW of farm animals is impractical since measuring more than one linear morphometric trait under farm conditions is time-consuming and increases measurement error. Therefore, the current study concurs with previous findings, confirming that estimating BW from a single linear morphometric trait, such as CG, will be a viable option when there is no measuring scale available in the farmer's conditions. Moreover, CG is a good predictor of BW due to its contribution to muscle and bone development (Okpeku *et al*, 2011). In addition, from a field point of view, CG is less affected by the posture and restraints of animals and is easy to measure with minimum error compared to other morphometric traits.

Canonical discriminant analysis is a statistical tool which has been increasingly used to reduce the dimensionality of data (Macena *et al*, 2024) and identifies the linear combination of multiple traits that provide maximum correlation between groups (Takele *et al*, 2021; Melesse *et al*, 2022). The goal of such analysis is to reduce within-population variation and maximize between-group variations, which allows understanding of genetic similarity and diversity of the studied animal populations. The values of the two CANs in the present study were comparable with the findings of Tade *et al* (2021) who reported that CAN1 and CAN2 explained 70% and 30% of total variation in south Gondar goats, respectively, in accordance with the current findings. The significant difference between the two canonical variables in this study aligns with previous reports on various indigenous goat populations (Takele *et al*, 2021; Melesse *et al*, 2022; Guyo *et al*, 2023). In contrast, Selolo *et al* (2015) found that only the first canonical variable was significant. CD and SW showed the highest loading values, making them key traits for differentiating goat populations. These observations are consistent with those reported by Al-Atiyat *et al* (2024). The value of canonical correlation (0.74) for CAN1 indicates the strength of the function to discriminate the goats reared in the three agroecological zones. This indicates that the variables associated with CAN1, such as CD and SW, are most important variables to effectively discriminate the studied goat populations. In line with this finding, this strong correlation of CAN1 was reported by Ofori *et al* (2021), Ali *et al* (2024), Takele *et al* (2021) and Guyo *et al* (2023).

The Mahalanobis distance expresses the magnitude of variations at the phenotypic level among animal breeds/populations or groups (Dauda *et al*, 2018; Maksimović *et al*, 2023). The observed significant Mahalanobis distance indicates that the goat populations reared in three agroecological zones are morphologically different. The largest distance between the highland and lowland goat populations might be associated with geographical isolations between these two goat populations and management differences practiced by the farming communities of different agroecological zones (Ofori *et al*, 2021). The shortest distance between the highland and midland goats may indicate that these goats are closely related in the expression of their morphometric traits, which might have resulted in admixture and gene flow in both directions. The observed

Mahalanobis distance values in this study closely agree with those reported for Boer and red Sokoto goats (Muhammad *et al*, 2021), while they were higher than those reported for Hararghe highland goats (Takele *et al*, 2021). On the other hand, higher values than those observed in the current study were reported for different indigenous goat breeds reared in Indonesia (Depison *et al*, 2020), Serbia (Maksimović *et al*, 2023) and Ghana (Ofori *et al*, 2021). The cluster analysis showed that the highland and midland goat populations were closely clustered, which might be due to their geographical proximity, leading to potential gene flow between these two goat populations. The Wilk's lambda is the proportion of total variability not explained by the discriminator variables between populations (Takele *et al*, 2021). The value of Wilk's lambda in the current study is 31%, which indicates that 69% of the variability in the discriminating traits was due to variation between populations rather than within populations. These results are comparable with those reported by Guyo *et al* (2023) for Arsi-Bale goats.

The quadratic discriminant analysis showed that 86% of the total goat populations were correctly assigned to their original source, suggesting the existence of more similarity within the goat population than across the populations. A higher success rate of classification was reported for different indigenous goats of Ethiopia (Zergaw *et al*, 2017; Melesse *et al*, 2022; Guyo *et al*, 2023). On the other hand, a lower classification rate was reported for Hararghe highland goats (Takele *et al*, 2021), South Gondar goats (Tade *et al*, 2021) and West African Dwarf goats (Ofori *et al*, 2021). The observed inconsistency between classification and cross-validation may suggest the existence of a relatively heterogeneous goat population as a result of admixture and migration of animals (Takele *et al*, 2021; Melesse *et al*, 2022). The higher misclassification rate observed for goat populations raised in the midland agroecological zones might have been caused by gene flow from goats of lowland and highland agroecological zones through marketing channels and migration due to various reasons (Belayhun *et al*, 2023).

Conclusion

This study showed the existence of significant morphometric diversity among indigenous goat populations reared across the three agroecological zones of southeastern Ethiopia. These findings provide a foundation for designing sustainable breeding and conservation strategies to retain the observed morphometric uniformity within the population of each agroecological zone. Further studies should integrate genomic tools to validate the observed phenotypic diversities among the three agroecological zones.

Supplemental data

Supplemental Table 1. Morphometric traits recoded

Supplemental Table 2. Univariate test statistics to identify the relative importance of each morphometric trait to discriminate the goats' population in the three agroecologies.

Supplemental Table 3. Total-sample standardized canonical coefficients.

Supplemental Table 4. Total canonical structure.

Supplemental Figure 1. Graphical presentation of the regression equation for the estimation of male and female goats using the combined data.

Acknowledgments

The authors are grateful to Hawassa University, Office of Vice President for Research and Technology Transfer (VPRTT) for the financial support to undertake the study. We are thankful to households and development agents who assisted during data collection.

Author contributions

All authors have contributed to the conception and design of the study. Kebede Tilahun participated in data collection, data analysis and interpretation, writing the first draft and revision. Aberra Melesse contributed data analysis and interpretation and revision. Simret Betsha contributed by reviewing and editing the manuscript. All authors have read and approved the manuscript.

Conflict of interest statement

The authors declare that they have no conflict of interest.

Ethics statement

The study was reviewed and approved by the Research Ethics Review Committee (RERC) of Hawassa University (Reference No: REC 016/23).

References

Abebe, T., Korato, M. (2020). Phenotypic characterization of indigenous goats in east Arsi Zone, Oromia regional state. Global Journal of Animal Scientific Research 8 (1), 75–92. url: http://www.gjasr.com/index.php/GJASR/article/view/36.

Akounda, B., Ouédraogo, D., Soudré, A., Burger, P.A., Rosen, B.D., Van Tassell, C.P., Sölkner, J. (2023). Morphometric characterization of local goat breeds in two agroecological zones of Burkina Faso, West Africa. Animals 13, 1931.doi: https://doi.org/10.3390/ani13121931

Al-Atiyat, R. M., Suliman, G. M., Abu-Alruz, K., Al-Zyoud, F., Mamkagh, A., El-Waziry, A. M., Khan, R. U. (2024). Multivariate discriminant analysis on differentiating sheep breeds based on live body and carcass measurements. Pakistan Journal of Zoology 56(2). doi: https://dx.doi.org/10.17582/journal.pjz/20230816022706

Ali, S.S., Kuralkar, S.V., Kanadkhedkar, H.L., Bankar, P. S., Kataria, R. S., Vohra, V. (2024). Discriminant analysis reveals differences in breed structure in river buffaloes of Maharashtra (India).Tropical Animal Health and Production 56, 38. doi: https://doi.org/10.1007/s11250-024-03885-4

Aliyu, A. M., Rotimi, E. A., Galadima, N. M. (2021). Effect of age and sex on morphometric measurements of Sahelian goats in Faskari local government area of Katsina state. FUDMA Journal of Sciences 5 (1), 11-15. doi: https://doi.org/10.33003/fjs-2021-0501-507

Arsoy, D., Abdourhamane, I. M., İsfendiyaroğlu, M. (2024). Phenotypic characterization of indigenous Cyprus native

hair goat breeds using quantitative and qualitative trait analysis. Revista Cientifica de la Facultade de Veterinaria, 34(1).

Belayhun, T., Getachew, T., Mustefa, A., Engdawork, A. (2023). Phenotypic characterization of goat populations of North Shoa and South Wollo zones in Amhara region. Ecological Genetics and Genomics 27,100169.doi: https://doi.org/10.1016/j.egg.2023.100169

Birhaniea, M., Alemayehu, K., Mekuriaw, G. (2019). Morphological characterization of goat populations in central zone of Tigray. Tropical Animal Science Journal 42(2), 81–89. doi: https://doi.org/10.5398/tasj.2019.42.2.81

Bousbia, A., Boudalia, S., Gueroui, Y., Hadded, K., Bouzaoui, A., Kiboub, D., Symeon, G. (2021). Use of multivariate analysis as a tool in the morphological characterization of the main indigenous bovine ecotypes in northeastern Algeria. PLoS ONE 16(7), e0255153. doi: https://doi.org/10.1371/journal.pone.0255153

Chebo, C., Melesse, A., Betsha, S. (2023). Morphological descriptors of indigenous chickens: as a selection guide. Journal of Applied Animal Research, 51(1), 729-742. doi: https://doi.org/10.1080/09712119.2023.2274956.

Chokoe, T. C., Matelele, T. C., Maqhashu, A., Ramukhithi, F. V., Mphahlele, T. D., Mpofu, T. J., Nephawe, K. A., Mtileni, B. (2020). Phenotypic diversity of South African indigenous goat population in selected rural areas. American Journal of Animal and Veterinary Sciences 15(1). doi: https://doi.org/10.3844/ajavsp.2020.59.66

Dauda, A., Abbaya, H.Y., Ebegbulem, V.N. (2018). Application of multifactorial discriminant analysis of morphostructural differentiation of sheep. Genetics and Genetic Engineering 2(2), 11-16.

Dea, D., Melesse, A., Mekasha, Y. (2019). Application of morphometric traits and body indices in assessing the type and function of local goats reared in two districts of Gamo-Gofa zone. Ethiopian Journal of Animal Production 19(1), 73–90.

Depison, D., Putra, W.P.B., Gushairiyanto, G., Alwi, Y., Suryani, H. (2020). Morphometric characterization of Kacang goats raised in lowland and highland areas of Jambi Province, Indonesia. Journal of Advanced Veterinary and Animal Research 7(4), 734–743. doi: https://doi.org/10.5455/javar.2020.g475

Deribe, B., Beyene, D., Dagne, K., Getachew, T., Gizaw, S., Abebe, A. (2021). Morphological diversity of north eastern fat-tailed and northwestern thin-tailed indigenous sheep breeds of Ethiopia. Heliyon, 7: 1-8. doi: https://doi.org/10.1016/j.heliyon.2021.e07472

Elzarei, M.F., Alhasyani, M.S., Aloufi, A.A., Mousa, E.F. (2023). Phenotypic characterization of indigenous sheep breeds in Saudi Arabia. Journal of Agricultural Science 15(8), 16–22. doi: https://doi.org/10.5539/jas.v15n8p16

FAO. (2012). Food and Agricultural Organization. Phenotypic characterization of animal genetic resources. url: https://www.fao.org/3/a1404e/a1404e.pdf.

Ganiyu, O., Osaiyuwu, O. H., Akinyemi, M. O., Salako, E. A. (2018). Genetic relationship among indigenous goat populations in Nigeria based on cellulose acetate electrophoresis systems. *Scientific Papers Animal Science and Biotechnologies* 51(2), 55-55.

Getahun, S., Ahmed, S., Zemene, W. (2020). Morphometric characterization of indigenous goats in East Gojjam zone, Amhara region, Ethiopia. *International Journal of Advanced Research in Biological Sciences 7*, 47-62. url: https://ijarbs.com/pdfcopy/2020/feb2020/ijarbs5.pdf

Getaneh, M., Taye, M., Kebede, D. (2022). Conformation trait characterization of indigenous goats in selected districts of East Gojjam zone, Amhara region, Ethiopia. *Journal of Applied Animal Research* 50(1), 225–238. doi: https://doi.org/10.1080/09712119.2022.2058516

Ghahri, B., Alijani, S., Rafat, S. A., Nabavi, R., Kia, H. D. (2019). Genetic appraisal of growth traits in Iranian native Ghezel sheep using random regression models. *Turkish Journal of Veterinary and Animal Sciences* 43(3), 372-379. doi: https://doi.org/10.3906/vet-1704-73

Guyo, M., Melesse, A., Taye, M. (2023). Multivariate analysis of morphometric traits revealed phenotypic diversity of Arsi-Bale goats reared in Bale eco-regions, Ethiopia. *Agriculture Conspectus Scientificus 88*(4), 325-335.url: https://hrcak.srce.hr/311921

Habimana, R., Ngeno, K., Mahoro, J., Ntawubizi, M., Shumbusho, F., Manzi, M., Okeno, T.O. (2021). Morpho-biometrical characteristics of indigenous chicken ecotype populations in Rwanda. *Tropical Animal Health and Production* 53, 24.doi: https://doi.org/10.1007/s11250-020-02475-4

Hankamo, A., Woldeyohannes, T., Banerjee, S. (2020). Morphometrical characterization and structural indices of indigenous goats reared in two production systems in Sidama zone, southern Ethiopia. *International Journal of Animal Science and Technology* 4(1), 6-16. doi: https://doi.org/10.11648/j.ijast.20200401.12

Kandoussi, A., Petit, D., Boujenane, I. (2021). Morphologic characterization of the Blanche de Montagne, an endemic sheep of the Atlas Mountains. *Tropical Animal Health and Production* 53, 154. doi: https://doi.org/10.1007/s11250-021-02577-7

Kitila, K.B., Aleli, A.T., Tufa, S.S. (2025). Morphological characterization of indigenous goats in selected districts of West Shewa zone, Oromia regional state, Ethiopia. PLoS One 20 (7): e0327309. doi: https://doi.org/10.1371/journal.pone.0327309

Kuthu, Z.H., Ahmed, Z., Nawaz, M., Bila, G., Zubair, M., Ahmed, I., Ikram, M. (2022). Principal component analysis of Kail sheep based on body measurements. *Pakistan Journal of Science* 74(4). doi: https://doi.org/10.57041/pjs.v74i4.801

Laouadi, M., Tennah, S. Moula, N., Antoine-Moussiaux, N., Kafidi, N. (2020). Morphological characterization of indigenous goats in the region of Laghouat in Algeria. *Archivos de Zootecnia* 69(267), 272–279.url: https://vufind-ceipa.metacatalogo.com/Record/ojs-dialnet-ART0001410169

Macena, E. C. B. da C., Costa, R. G., de Sousa, W. H., Cartaxo, F. Q., Arandas, J. K. G., Ribeiro, M. N., Ribeiro, N. L. (2024). Canonical discriminant analysis on the characterization of the goat carcass. *Acta Scientiarum-Animal Sciences 46*, 1–7. doi: https://doi.org/10.4025/actascianimsci.v46i1.58440

Maksimović, N., Cekić, B., Ćosić, I., Ružić Muslić, D., Caro Petrović, V., Stojiljković, N., Stanišić, N. (2023). Discriminant analysis approach in morphometric differentiation and characterization of Serbian Autochthonous goats. *Animals 13*(12), 1952. doi: https://doi.org/10.3390/ani13121952

Mdladla, K., Dzomba, E.F., Muchadeyi, F.C. (2017). Characterization of the village goat production systems in the rural communities of the Eastern Cape, KwaZulu-Natal, Limpopo and North West Provinces of South Africa. *Tropical Animal Health and Production*. 49, 515–527. doi: https://doi.org/10.1007/s11250-017-1223-x

Mebratie, W., Tekuar, S., Alemayehu, K., Dessie, T. (2022).

Body weight and linear body measurements of indigenous goat population in Awi Zone, Amhara region, Ethiopia. *Acta Agriculturae Scandinavica A: Animal Sciences* 71(1–4), 89–97.doi: https://doi.org/10.1080/09064702.2022.2127868

Melesse, A., Yemane, G., Tade, B., Dea, D., Kayamo, K., Abera, G., Mekasha, Y., Betsha, S., Taye, M. (2022). Morphological characterization of indigenous goat population in Ethiopia using canonical discriminant analysis. *Small Ruminant Research*, *206*.106591. doi: https://doi.org/10.1016/j.smallrumres.2021.106591

Monau, P.I., Visser, C., Nsoso, S.J. (2018). Phenotypic and genetic characterization of indigenous Tswana goats. *South African Journal of Animal Science 48*(5).doi: http://dx.doi.org/10.4314/sajas.v48i5.12

Muhammad, H.A., Garba, Y., Ogah, D.M. (2021). Multivariate analysis of morphometric differentiation in the Red Sokoto and Boer Goats. *Advances in Animal and Veterinary Sciences* 9(4),595–603.

Muluneh, T., Tadesse, W. (2022). Phenotypic characterization of indigenous goat population in southern, Ethiopia. *American Journal of Life Sciences* 10(3), 31. doi: https://doi.org/10.11648/j.ajls.20221003.11

Nantongo, Z., Agaba, M., Shirima, G., Mugerwa, S., Opiyo, S., Mrode, R., Munishi, L. (2024). Variability in body weight and morphology of Uganda's indigenous goat breeds across agroecological zones. *PLOS ONE*, *19*(1), e0296353. doi: https://doi.org/10.1371/journal.pone.0296353

Nguluma, A., Kyallo, M., Tarekegn, G. M., Loina, R., Nziku, Z., Chenyambuga, S., Pelle, R. (2022). Typology and characteristics of indigenous goats and production systems in different agro-ecological zones of Tanzania. Tropical Animal Health and Production, 54(1), 70. doi: https://doi.org/10.1007/s11250-022-03074-1 peter

Ntonga, S.P., Akinmoladun, O.F., Mpetile, Z. (2025). Phenotypic characterization of indigenous Xhosa goat ecotype in three agro-ecological zones in the Eastern Cape province, South Africa. Veterinary and Animal Science 30, 100512. doi: https://doi.org/10.1016/j.vas.2025.100512

Ofori, S. A., Hagan, J. K., Kyei, F. (2021). Morphometric characterization and differentiation of West African Dwarf goat populations in Ghana. *Tropical Animal Health and Production* 53,69. doi: https://doi.org/10.1007/s11250-020-02498-x

Okoro, V. M. O., Chikereuba, U., Iheanacho, G., Ogundu, A., Chukwuocha, C., Chikaodinaka, F., Ifediba, C., Kadurumba, O. E., Ogundu, U., Iloeje, M. (2023). West African Dwarf goat populations of tropical rainforest zone: application of multivariate analysis tools on their characterization using morpho-biometric traits. *Tropical Animal Health and Production*, 55(3), 1–24. doi: https://doi.org/10.1007/s11250-023-03641-0

Okpeku, M., Yakubu, A., Peters, S. O., Ozoje, M. O., IKEOBI, C. O. N., Adebambo, O. A., Imumorin, I. G. (2011). Application of multivariate principal component analysis to morphological characterization of indigenous goats in Southern Nigeria. *Acta Agriculturae Slovenica*, *98*(2), 101-109.

Oladepo, A.D., Salako, A.E. (2017). Blood protein polymorphism and differentiation of selected indigenous cattle breeds in Nigeria. Nigerian Journal of Animal Production. 44:11–18.

Razgour, O., Forester, B., Taggart, J. B., Bekaert, M., Juste, J., Ibáñez, C., Manel, S. (2019). Considering adaptive genetic variation in climate change vulnerability assessment reduces species range loss projections. *Proceedings of the National Academy of Sciences*, *116*(21), 10418-10423.

Rotimi, E.A., Momoh, O.M., Egahi, J.O., Dim, N.I. (2020). Multivariate principal component analysis of morphological traits in West African Dwarf goats. *FUDMA Journal of Sience 4*(1), 333 – 339.

Sako, T., Tyasi, T. L, Ng'ambi, J. (2024). Phenotypic characterization of selected African sheep breeds: A Review. *Pakistan Journal of Zoology* 1-13. doi: https://dx.doi.org/10.17582/journal.pjz/20230510000554

Sarma, L., Nahardeka, N., Goswami, R. N., Aziz, A., Zaman, G., Das, A., Akhtar, F. (2019). Non-genetic factors affecting pre-weaning growth and morphometric traits in Assam Hill goat. *Veterinary World* 12(8), 1327–1331.

SAS (Statistical Analysis System). 2016. SAS for Windows, ver. 9.4 SAS Institute, Inc., Cary, NC, USA.

Selolo, T.C., Mashiloane, M. L., Norris, D., Ng'ambi, J.W., Brown, D. (2015). Morphological differentiation of indigenous goats in different agroecological zones of vhembe district, Limpopo province, South Africa. *Indian Journal of Animal Research 49*(4), 527–531. doi: https://doi.org/10.5958/0976-0555.2015.00052.7

Selvan, A.S., Thangaraj, S., Dash, S., Karthikeyan, A., Karthickeyan, S. M. K. (2023). Multivariate analysis of morphometric traits of Malaimadu cattle—autochthonous draft cattle of south India. *Tropical Animal Health and Production 55*(6), 369. doi: https://doi.org/10.1007/s11250-023-03790-2

Sheriff, O., Alemayehu, K., Haile, A., (2021). Phenotypic ranking experiments in identifying breeding objective traits of smallholder farmers in northwestern Ethiopia. *PLOS ONE* 16 (3): e0248779. doi: https://doi.org/10.1371/journal.pone.0248779

Shuma, G., Mekonnen, A., Duguma, G. (2024). Phenotypic characterization of indigenous goat types in Kellem Wollega zone of Oromia Region, Ethiopia. Ethiopian Journal of Agricultural Science 34(3) 34-56.

Tade, B., Melesse, A., Betsha, S. (2021). Application of multivariate analysis for the differentiation of indigenous goat populations of South Gondar, Ethiopia. *Journal of Agricuture and Environmental Siences 6*(1), 19–31.

Takele, A., Melesse, A., Taye, M. (2021). Application of multivariate analysis to differentiate Hararghe Highland goat populations reared in the west Hararghe zone, Ethiopia. *Chilean Journal of Agricultural and Animal Sciences* 37(3), 209–220. doi:https://doi.org/10.29393/CHJAAS37-23AMAM30023

Tarekegn, G. M., Khayatzadeh, N., Liu, B., Osama, S., Haile, A., Rischkowsky, B., Mwacharo, J. M. (2021). Ethiopian indigenous goats offer insights into past and recent demographic dynamics and local adaptation in sub-Saharan African goats. *Evolutionary applications* 14(7), 1716-1731. doi: https://doi.org/10.1111/eva.13118

Tenagne, A., Taye, M., Dessie, T., Muluneh, B., Kebede, D., Tarekegn, G.M. (2023). Quantifying morphometric and adaptive characteristics of indigenous cattle genetic resources in northwest Ethiopia. *PLOS ONE* 18(3): e0280640. doi: https://doi.org/10.1371/journal.pone.0280640

Tyasi, T.L., Tada, O. (2023). Principal component analysis of morphometric traits and body indices in South African Kalahari Red goats. *South African Journal of Animal Science* 53(1).doi: https://doi.org/10.4314/sajas.v53i1.04

Valsalan, J., Sadan, T., Venketachalapathy, T. (2020). Multivariate principal component analysis to evaluate growth performances in Malabari goats of India. *Tropical Animal Health and Production* 52 (5), 2451–2460 .doi:

https://doi.org/10.1007/s11250-020-02268-9
Vanvanhossou, S. F.U., Diogo, R.V.C., Dossa, L.H. (2018). Estimation of live bodyweight from linear body measurements and body condition score in the West African Savannah shorthorn cattle in North-West Benin. *Cogent Food and Agriculture* 4(1), 1549767. doi: https://doi.org/10.1080/23311932.2018.1549767
Velado-Alonso, E., Morales-Castilla, I., Gómez-Sal, A. (2020). Recent land use and management changes decouple the adaptation of livestock diversity to the environment. *Scientific Reports*, *10*(1), 21035.
Whannou, H.R., Afatondji, C.U., Ahozonlin, M.C., Spanoghe, M., Lanterbecq, D., Demblon, D., et al. (2021) Morphological variability within the indigenous sheep population of Benin. *PLOS ONE* 16 (10): e0258761.doi: https://doi.org/10.1371/journal.pone.0258761
Worku, A., Melesse, A. (2021). Application of multivariate analysis on morphometric traits to characterize the sheep populations reared in the central rift valley of Ethiopia. *Ethiopian Journal of Animal Production* 21(1), 30–47.
Yamane, T. (1967). Statistics: An Introductory Analysis, 2nd edition, New York: Harper and Row.
Zergaw, N., Dessie, T., Kebede, K. (2017). Using morphometric traits for live body weight estimation and multivariate analysis in Central Highland and Woyto-Guji goat breeds, Ethiopia. *African Journal of Agricultural Research 12*(15),1326–1331. doi: https://doi.org/10.5897/ajar2016.11628

Genetic Resources

ORIGINAL ARTICLE

Genetic Resources (2026), 7 (13), 41–56
DOI: 10.46265/genresj.TDIC2170
https://www.genresj.org
ISSN: 2708-3764

Phenotypic variability of tarwi (*Lupinus mutabilis* S.) in Peruvian germplasm collections

Kevin Ortega-Quispe[a,b,*], Eunice Peña-Elme[a], Carolina Girón-Aguilar[a], Nery Amaro-Camarena[a], Claudia Rios-Chavarría[c], Bertha Lopez-Pariona[a], Francis Cerrón-Mercado[a], Steve Camargo-Hinostroza[b], Samuel Pizarro[d]

[a] *Estación Experimental Agraria Santa Ana, Dirección de Recursos Genéticos y Biotecnología, Instituto Nacional de Innovación Agraria (INIA), Carretera Saños Grande – Hualahoyo Km 8 Santa Ana, Huancayo, Junín 12007, Peru*
[b] *Escuela Académico Profesional de Ingeniería Ambiental, Universidad Continental, Huancayo 12001, Peru*
[c] *Faculty of Forestry, University of Sopron, Bajcsy-Zsilinszky u. 4, 9400, Sopron, Hungary*
[d] *Centro de Investigación en Geomática Ambiental (CIGA), Instituto de Investigación para el Desarrollo Sustentable de Ceja de Selva (INDES-CES), Universidad Nacional Toribio Rodríguez de Mendoza de Amazonas, Chachapoyas 01001, Peru.*

Abstract: The growing global loss of genetic diversity, phenotypic characterization becomes essential for identifying resilient varieties capable of diversifying and strengthening the agricultural production of underutilized crops such as tarwi (*Lupinus mutabilis* S.). This study aimed to characterize the phenotypic variability of 41 tarwi accessions conserved in the germplasm bank of the National Institute of Agricultural Innovation (INIA) of Peru. The accessions were evaluated over two consecutive agricultural seasons at the Santa Ana Agrarian Experimental Station under local conditions. Thirty morphological descriptors (17 qualitative and 13 quantitative) were used following IBPGR guidelines. Data were analyzed using descriptive statistics, principal component analysis, hierarchical clustering and correlation analysis for quantitative descriptors, as well as frequency tables and the Shannon-Weaver diversity index for qualitative descriptors. The results revealed high phenotypic variability, particularly in traits related to yield, plant architecture and floral attributes. The accessions were grouped into three morpho-agronomic types: (1) highly productive accessions, (2) accessions with vigorous vegetative development, and (3) short-cycle plants with moderate yields. Yield per plant was significantly associated with the total pod number, total seed mass in hundred seeds and seed thickness. The study revealed considerable phenotypic diversity, characterized by significant correlations among key agronomic traits, the delineation of three distinct phenotypic clusters, and the identification of valuable qualitative attributes, which reinforces their potential for conservation and breeding programmes. However, expanded germplasm evaluation and multi-environment trials are required to validate genotype stability and refine selection criteria. However, additional accessions and further analyses are needed to validate the observed patterns.

Keywords: Germplasm, accessions, conservation, Leguminosae, descriptors

Citation: Ortega-Quispe, K., Peña-Elme, E., Girón-Aguilar, C., Amaro-Camarena, N., Rios-Chavarría, C., Lopez-Pariona, B., Cerrón-Mercado, F., Camargo-Hinostroza, S. and Pizarro, S. (2026) "Phenotypic variability of tarwi (*Lupinus mutabilis* S.) in Peruvian germplasm collections", *Genetic Resources*, 7(13), pp. 41–56. doi: 10.46265/genresj.TDIC2170.

Introduction

Agriculture and global food security faces unprecedented challenges due to climate change (Morales-Casco & Zúniga-González, 2016), environmental degradation (Palacios-López, 2024) and the growing dependence on a limited number of commercial crops for food, mainly staple cereals such as wheat, rice and maize, which has led to the neglect of a wide range of genetic resources and potentially valuable agronomic traits (Massawe *et al*, 2015). According to the Food and Agriculture Organization of the UN (FAO), more than 75% of agricultural genetic diversity has been lost over the past century, thereby increasing the vulnerability of agri-food systems to pests, diseases and extreme weather events (FAO, 2019). In this context, the agromorphological characterization of underutilized crops, such as Andean legumes, becomes an essential strategy to identify resilient varieties that can diversify agricultural production and

* Corresponding author: Kevin Ortega-Quispe (kevinorqu@gmail.com)

Received: 05.08.2025 | Accepted: 07.01.2026 | Published online: 28.01.2026

improve nutrition (Ojuederie *et al*, 2023).

The Andean region is considered one of the most important centres of origin and diversification of crops worldwide, due to its climatic and cultural variability (Flores *et al*, 2003). Over time, traditional crops such as tarwi (*Lupinus mutabilis* S.), which for centuries played a central role in Andean agriculture, have been gradually displaced by introduced species of greater commercial interest, such as faba bean (*Vicia faba* L.). This shift is driven not only by economic dynamics but also by sociocultural factors, as tarwi has historically been undervalued and associated with social stigmas often perceived as 'poor people's food' or 'indigenous food', which contributed to its progressive abandonment in agricultural systems (Chalampuente *et al*, 2021).

Lupinus mutabilis S. is a leguminous species of the Fabaceae family native to the Andes and is known by several vernacular names reflecting the linguistic diversity of its distribution range: 'chocho' (Colombia, Ecuador, northern Peru), 'tarwi' or 'tarhui' in Quechua (central and south-central Peru), 'tauri' in Aymara (south of Lake Titicaca, Peru and Bolivia), and 'chuchus muti' (Cochabamba, Bolivia) (Tapia, 2015). The species exhibits outstanding adaptability to diverse geographic regions, demonstrating promising agronomic performance and genetic variability not only across South American ecosystems but also in Mediterranean climates in Europe, where its cultivation potential has been successfully evaluated (Guilengue *et al*, 2019). Traditionally cultivated in small-scale farming systems, it stands out for its high nutritional value, with a protein content exceeding 40%, along with essential oils and key minerals important for human nutrition. In recent years, it has been recognized as a biofortified food due to the notable concentrations of micronutrients such as iron, zinc and boron present in its seeds. These attributes position it as a strategic crop in addressing food security and nutrition challenges in rural areas. Furthermore, the flour obtained from debittered tarwi exhibits favourable physicochemical properties: 9.20% moisture, 2.03% ash, 20.01% fat, 51.65% protein, 9.04% fiber and 8.10% carbohydrates, reinforcing its potential as a functional ingredient for the development of value-added food products (Caligari *et al*, 2000; Enrique Quispe, 2022; Vera-Vega *et al*, 2022), Additionally, through its nitrogen-fixing based on symbiosis with *Rhizobium* spp., combined with high tolerance to poor soils and optimal adaptation to elevations between 2,000 and 3,800m a.s.l. in temperate to cold climates, tarwi contributes significantly to soil improvement (Tapia, 2015).

Several studies in countries such as Ecuador and Bolivia have highlighted the notable phenotypic plasticity of tarwi (Camarena *et al*, 2012; Peralta *et al*, 2013; Aguilar Angulo, 2015; Cano *et al*, 2022). Nonetheless, in Peru, systematic comparisons of accessions from different regions remain lacking. Although novel debittering techniques have been developed, most efforts have focused on reducing alkaloid content or improving productivity, without integrating the selection of high-yielding genotypes, limiting the effectiveness of genetic improvement programmes (Gulisano *et al*, 2022). This gap underscores the current weaknesses and the urgent need to conserve and characterize tarwi's genetic diversity within germplasm banks, as a strategic foundation for the development of sustainable and nutritious cropping systems that contribute to food security. Given the increasing global demand for sustainable plant-based foods, it is essential to prioritize the generation of standardized data for the description and comparison of germplasm collections, in order to promote the conservation and use of underutilized crops like tarwi within food sovereignty strategies (FAO, 2010; Padulosi *et al*, 2011).

Additionally, there is a need to provide farmers and stakeholders with agromorphological information on well-adapted varieties and accessions to reduce production risks, inform public food security policies, and establish a replicable methodology based on standardized qualitative and quantitative descriptors by the International Board for Plant Genetic Resources (IBPGR, 1981). This information aims to complement the global efforts of germplasm banks and research institutes for the conservation and improvement of plant genetic resources (Gresta *et al*, 2017), positioning tarwi as a sustainable alternative for high-Andean agri-food systems.

Accordingly, this research aimed to characterize the phenotypic diversity of 41 tarwi accessions from the Germplasm Bank of the Peruvian National Institute of Agrarian Innovation (INIA), using descriptors established by IBPGR protocols, evaluated over two consecutive seasons under the agroecological conditions of the central highlands at the Santa Ana Agricultural Experiment Station (hereafter EEA Santa Ana). Ultimately, the study seeks to expand current knowledge on tarwi morphology and generate a positive impact on the agricultural sector by promoting the conservation of plant genetic resources and supporting progress toward the UN Sustainable Development Goals (SDGs) (Seyedsayamdost, 2020).

Materials and methods

Plant material

The research was carried out using a total of 41 tarwi accessions, which are part of the INIA Germplasm Bank. The accessions were originally collected across diverse regions of Peru during 1979 and 1981 and have since been maintained through periodic regeneration following protocols established by the Sub-Directorate of Genetic Resources, contributing to national conservation efforts for Andean crop genetic diversity.

The collection sites, number of accessions, and their altitudinal range of origin were determined based on passport data recorded in the national tarwi collection (Table 1), where each accession has a unique international identification code 'PER'. The full information on the tarwi accessions, as well as the base material used in this study, is available on Zenodo (https://doi.org/10.5281/zenodo.15740283).

Study area

The study was conducted at the EEA Santa Ana, located in the district of El Tambo, province of Huancayo, Junín region (12°00'42.36" S, 75°13'17.60" W), in the southeastern part of the Mantaro Valley, an important agricultural area in the Peruvian Andes (Figure 1). The station is situated between 3,303 and 3,325m a.s.l., on an alluvial fan with predominantly flat topography (Enriquez *et al*., 2025).

During the evaluation period (2022–2024), temperatures ranged from 5.56°C to 22.42°C; relative humidity varied between 60.35% and 86.07%, while monthly precipitation fluctuated between 11.8mm (October 2024) and 145.9mm (February 2024), according to data from the National Meteorology and Hydrology Service (SENAMHI, 2024).

Table 1. List of tarwi accessions evaluated in the 2022–2023 and 2023–2024 agricultural campaigns.

Department of Peru	Code	# of Accessions	%	Altitudinal range (m a.s.l)
Amazonas	PER1006833	1	2.43	2,770
Apurímac	PER1006006	1	2.43	3,500
Cajamarca	PER1006664; PER1006673 PER1006676; PER1010656 PER1010657; PER1010658 PER1006722; PER1010659 PER1006734; PER1006807 PER1010661; PER1010662 PER1010663; PER1010664 PER1006827; PER1010666 PER1006921; PER1006922 PER1006934; PER1010667 PER1006116; PER1006157 PER1006216; PER1006223	20	48.78	2,600–3,460
Cusco	PER1006276; PER1006299 PER1006310; PER1006382 PER1010654	9	21.95	2,651–3,391
Junín	PER1006446; PER1006495	2	4.87	3,209–3,507
La Libertad	PER1006523; PER1006549 PER1006811; PER1006815	4	6.98	2,600–4,000
Puno	PER1006555; PER1006570 PER1006571; PER1006590	4	11.63	3,035–3,895
Total accessions	-	41	100	-

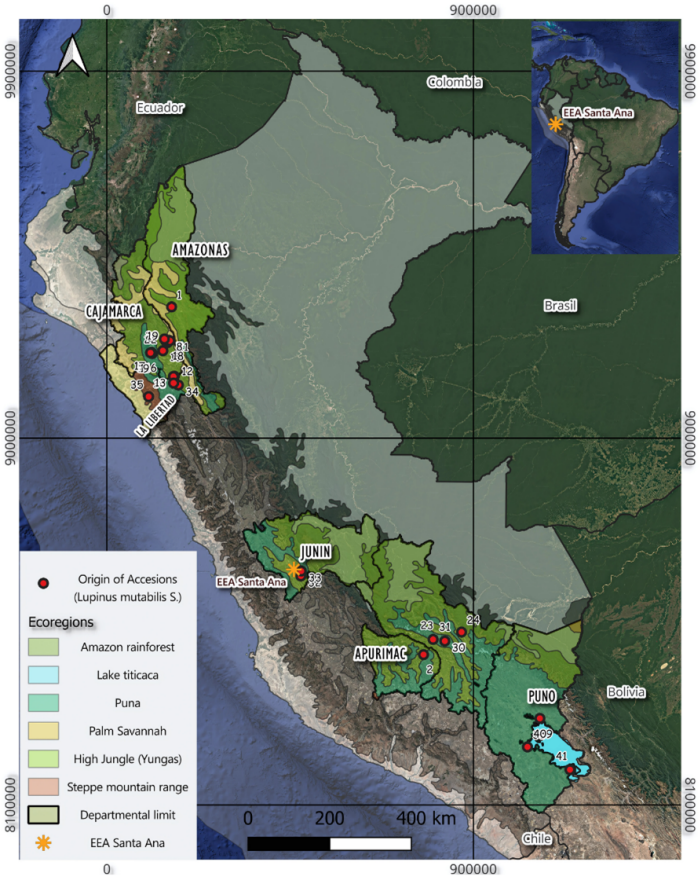


Figure 1. Regions of origin of the 41 tarwi accessions in Peru and location of the EEA Santa Ana.

The research was carried out in plots No. 22 and 24 of EEA Santa Ana over two consecutive growing seasons (2022–2023 and 2023–2024), located at an average altitude of 3,295m a.s.l. (12°00'16" S, 75°13'16" W). Sowing took place on clay-loam soils with slightly acidic pH (6.2–6.6), electrical conductivity ranging from 2.4 to 5.4mS/m, and organic matter content between 1.5% and 3.09%, under controlled rigation conditions.

Methodology

Soil preparation and cultivation of plants

Land preparation in both growing seasons involved the formation of four furrows per plot, each 5m in length. Each plot was sown with 96 seeds per accession, placing three seeds per planting site, establishing a spatial arrangement of 0.80m between plants and 1.0m between rows. Two-meter-wide alleys were left between plots, in which maize (*Zea mays*) was planted to prevent cross-pollination, facilitate evaluations and ensure the genetic representativeness of each accession.

Regarding cultural or agronomic management, seeds were disinfected with a 0.5% solution of Vitamax 300 (Carboxin 200g/Kg + Captan 200g/Kg) to prevent disease. Manual weeding was carried out between 35 and 45 days after sowing. To ensure plant establishment, four flood irrigations were applied, and soil moisture was monitored in later stages, taking advantage of seasonal rainfall from December to May.

Fertilization was applied at two stages: 50% at sowing and the remaining 50% during hilling. The nutrient ratio used was 20-80-40 of N, P_2O_5 and K_2O, ensuring adequate nutrient supply for optimal crop development. Routine weeding and irrigation were carried out as needed throughout the season.

A total of 41 plots, each measuring 20m² and consisting of four rows, were established and evaluated. During each growing season, five plants were randomly selected from the central rows of each plot to minimize edge effects, individually labelled, and subjected to agromorphological characterization. The total area allocated to the experiment was approximately 1,000m² per agricultural season.

Morphological characterization

A total of 30 agromorphological descriptors were evaluated; of these, 26 were adopted from IBPGR (1981) as reference, while the remaining 4 were proposed by the researchers based on local agroecological conditions and the priorities of the conservation and breeding programme. Of the 30 descriptors, 17 were qualitative and 13 quantitative. During flowering, evaluations were conducted when more than 50% of the plants per accession displayed flowers. The intensity of flower colour was observed, and a general qualitative assessment was recorded through observation of the entire plot. Days to harvest were evaluated as the number of days from sowing until 50% of the pods per accession reached physiological maturity and were suitable for harvest.

Qualitative descriptors

Table 2 presents the qualitative descriptors used in the agromorphological characterization of tarwi throughout the crop's phenological cycle. Evaluation considered various plant structures, including morphological characteristics of the stem, leaves, flowers, inflorescences, pods, and seeds.

Quantitative descriptors

For the agromorphological characterization, ten quantitative descriptors were used, specifically selected according to the guidelines established by IBPGR (1981) for lupins (Table 3), of which three descriptors were proposed by researchers from EEA Santa Ana. These descriptors made it possible to evaluate phenotypic variations related to plant development and morphology, such as yield per plant, yield per plot, days to harvest, among others, contributing to a better understanding of the genetic diversity present in the evaluated material.

Field evaluations

Field observations began during the seedling stage, ten days after sowing (DAS), once at least 50% of the cotyledons had emerged in an accession. During the flowering stage (110 to 150 DAS), evaluations were conducted when 50% of the plants had flowers. Traits recorded included stem thickness, number of primary branches, number of leaflets per leaf, and days to reach 50% flowering. For the flowers, the colour and intensity of the floral bud before opening were assessed, as well as the colour and intensity of the wings and keel of the newly opened flower. The colouration of the marginal band, central spot and intermediate zone of the standard petal was also described, both in the open flower and before wilting. In addition, the length of the main inflorescence was measured.

The maturation phase was evaluated when more than 50% of the plants showed mature pods. At this stage, the colour and intensity during pod formation and maturity (green stage), presence of pubescence on mature pods, number of pods on the main axis and per plant, as well as pod dimensions (length and width) were recorded.

Finally, seed evaluation was conducted after harvest (210–295 DAS), assessing seed shape, dimensions (length, width, thickness), brightness, colour and its intensity, weight of 100 seeds (with moisture content below 12%), and yield per plant and plot.

Statistical analysis

Statistical analyses were performed using combined data from two consecutive agricultural seasons (2022–2023 and 2023–2024), processed with R software version 4.3.3 (R Core Team, 2023). For the quantitative descriptors, a multivariate approach was applied in four stages. First, descriptive statistics (mean, range, standard deviation and coefficient of variation) were calculated to assess phenotypic variability among accessions. Second, a principal component analysis (PCA) was conducted using the FactoMineR package (Lê *et al*, 2008) and factoextra (Kassambara & Mundt, 2020), to reduce the dimensionality of the dataset and visualize the multivariate distribution of the accessions.

Based on the principal components, a hierarchical cluster analysis (HCA) was carried out using Euclidean distance and Ward's method (Everitt *et al*, 2011), allowing the identification of homogeneous morphological groups, which were represented by a dendrogram. A heatmap was then generated to simultaneously visualize the similarity between

Table 2. Qualitative morphological traits evaluated in the study, with their respective descriptors and categories according to the scoring scale. *, Variables proposed by the researchers of EEA Santa Ana.

Acronym	Morphological descriptor	Categories
PGPE*	Pigmentation of petioles	0 Absent, 1 Present
IFBCJBO	Intensity of flower bud colour just before opening	1 Very Light*, 3 Light, 5 Medium, 7 Dark, 9 Very dark*
FWCJBO	Flower wing colour just before opening	1 White, 2 Yellow, 3 Orange, 4 Pink, 5 Red, 6 Green, 7 Blue, 8 Purple, 9 Brown, 10 Lilac*
CCSSJOF	Colour of central spots of standard of just opened flower	0 Absent central spots, 1 White, 2 Yellow, 3 Orange, 4 Pink, 5 Red, 6 Green, 7 Blue, 8 Purple, 9 Brown, 10 Lilac*
IRCSJOF	Intermediate region colour of standard of just opened flower	0 Absent intermediate region, 1 White, 2 Yellow, 3 Orange, 4 Pink, 5 Red, 6 Green, 7 Blue, 8 Purple, 9 Brown, 10 Lilac*
IFKCJBO	Intensity of flower keel colour just before opening	1 Very light*, 3 Light, 5 Medium, 7 Dark, 9 Very dark*
FWCJBW	Flower wing colour just before wilting	1 White, 2 Yellow, 3 Orange, 4 Pink, 5 Red, 6 Green, 7 Blue, 8 Purple, 9 Brown, 10 Lilac*
MBCSFJBW	Marginal band colour of standard of flower just before wilting	0 Absent marginal band, 1 White, 2 Yellow, 3 Orange, 4 Pink, 5 Red, 6 Green, 7 Blue, 8 Purple, 9 Brown, 10 Lilac*
CCSSFJBW	Colour of central spots of standard of flower just before wilting	0 Absent central spots, 1 White, 2 Yellow, 3 Orange, 4 Pink, 5 Red, 6 Green, 7 Blue, 8 Purple, 9 Brown, 10 Lilac*
IRCSFJBW	Intermediate region colour of standard of flower just before wilting	0 Absent intermediate region, 1 White, 2 Yellow, 3 Orange, 4 Pink, 5 Red, 6 Green, 7 Blue, 8 Purple, 9 Brown, 10 Lilac*
FKCJBW	Flower keel colour just before wilting	1 White, 2 Yellow, 3 Orange, 4 Pink, 5 Red, 6 Green, 7 Blue, 8 Purple, 9 Brown, 10 Lilac*
UFP	Uniformity of flowering of the plot	0 None, 3 Little, 5 Medium, 7 Very Much
MPP	Mature pod pubescence	0 Absent, 3 Slight, 5 Medium, 7 Strong
SSH	Seed shape	1 Spherical, 2 Flattened spherical or lenticular, 3 Oval, 4 Flattened oval, 5 Cuboid, 6 Flattened cuboid
SL	Seed luster	1 Mate, 2 Brilliant
PSC	Primary seed colour	1 White, 2 Yellow, 3 Orange, 4 Pink, 5 Red, 6 Green, 7 Blue, 8 Purple, 9 Brown, 10 Lead*, 11 Bayo*, 12 Black*
IPSC	Intensity primary seed colour	1 Very Light*, 3 Light, 5 Medium, 7 Dark, 9 Very Dark*

Table 3. Quantitative morphological characteristics observed in the study, descriptors and unit of measurement. *, Variables proposed by the researchers of EEA Santa Ana.

Acronym	Morphological descriptor	Unit
ST	Stem thickness	mm
NPB	Number of primary branches	Number branches per plant
LNS	Leaflets number per leaf	Number
LI	Length of inflorescences	cm
NPCA	Number of pods per central axis	Number per central axis
TNPP	Total number of pods per plant	Number per plant
PLMS	Pod length on the main stem	cm
PWMS	Pod width on the main stem	cm
TSWHS	Total seed mass in hundred seeds	g
YPLA	Yield per plant	g
YPLO*	Yield per plot	kg/ha
STH*	Seed thickness	mm
DH*	Days to harvest	days

accessions and the correlations among variables, facilitating the exploration of phenotypic variation patterns (Wilkinson & Friendly, 2009). Finally, Pearson correlation analysis (Pearson, 1895) was applied to assess associations among structural, reproductive and productive traits, identifying significant relationships between key variables.

For the analysis of qualitative descriptors, the most representative data per plot were considered. A frequency analysis was conducted, and the Shannon-Weaver diversity index (H′) was incorporated to evaluate the phenotypic diversity of each trait. The standardized value of H′ was calculated using its corresponding formula:

$$H' = -\sum_{i=1}^{n} P_i \ln(P_i)$$

The Shannon-Weaver diversity index (H′) is a metric that helps to understand the variety of traits within a group. Its calculation requires *n*, the total number of different categories or traits observed for a specific descriptor, and P_i, which represents the relative proportion of individuals exhibiting a particular trait. The calculation also involves *ln*, the natural logarithm.

According to Eticha *et al* (2005), the index values are interpreted as follows: low diversity ($0.10 \leq H \leq 0.40$), intermediate diversity ($0.40 \leq H \leq 0.60$), and high diversity ($H \geq 0.60$).

Results and discussion

Quantitative variables of the tarwi collection

Descriptive analysis and Shannon-Weaver diversity index

The descriptive analysis of 13 quantitative variables revealed notable genetic variability among the tarwi accessions (Table 4), as evidenced by the wide range of observed values. Yield per plant (YPLA) stood out, with a broad range from 22 to 276g and an average of 40.63 ± 2.6g/plant. This variation suggests a strong genetic component, with at least seven accessions exceeding 100g/plant, consistent with findings reported by Mujica *et al* (2021).

Yield per plot (YPLO) also showed high variability, ranging from 65.5 to 1697.5kg/ha. Among the reproductive traits, the total number of pods per plant (TNPP) showed the greatest dispersion, followed by the number of pods on the central axis (NPCA) and the number of primary branches (NPB), reflecting significant structural differences between accessions. In contrast, pod length (PLMS) and pod width (PWMS) were more stable, with averages of 9.48cm and 1.59cm, respectively, values that fall within the ranges reported by Gulisano *et al* (2019) for tarwi accessions evaluated under Andean conditions.

Other traits, such as inflorescence length (LI), number of leaflets per leaf (LNS) and stem thickness (ST), exhibited moderate variation. The most stable variables were days to harvest (DH) and seed thickness (STH), suggesting stronger genetic regulation.

Correlation analysis

Figure 2 illustrates the interrelationships among various agromorphological traits, highlighting particularly strong associations. A notable example is the high correlation observed between the number of primary branches (NPB) and the number of leaflets per leaf (LNS, r = 0.84), suggesting that plants with greater branching tend to develop more abundant foliage. This trait indicates a more vigorous and well-developed plant architecture, an essential aspect that previous studies on the diversity and agronomic performance of tarwi have positively linked to the crop's productive potential (Gulisano *et al*, 2022).

Another important finding is that yield per plant (YPLA) is significantly associated with reproductive traits such as the total number of pods (TNPP, r = 0.47), 100-seed weight (TSWHS, r = 0.45), and seed thickness (STH, r = 0.45). In other words, plants that produce more and better seeds tend to yield more, which is logical and has been confirmed in other studies on legumes (Saxena, 2018).

Table 4. Descriptive statistics and analysis of 13 quantitative morphological traits of tarwi accessions (Grouped values from 2022-2024).

Acronym	Descriptor	Min	Max	Range	Mean ± SE
DH	Days to harvest	211	295	84	259.8 ± 2.21
LI	Length of inflorescences	13	65	52	36.74 ± 0.41
LNS	Leaflets number per leaf	7	10	3	8.42 ± 0.04
NPB	Number of primary branches	4	27	23	15.4 ± 0.21
NPCA	Number of pods per central axis	6	34	28	16.3 ± 0.2
PLMS	Pod length on the main stem	5.92	12	6.08	9.48 ± 0.04
PWMS	Pod width on the main stem	1.2	2.2	1	1.59 ± 0.01
ST	Stem thickness	4.35	11.05	6.7	7.09 ± 0.05
STH	Seed thickness	3.69	8.31	4.62	5.33 ± 0.03
TNPP	Number of pods per plant	10	280	270	68.93 ± 1.87
TSWHS	Total seed mass in hundred seeds	21.84	39.82	17.98	28.58 ± 0.44
YPLA	Yield per plant	22	276	254	40.63 ± 2.6
YPLO	Yield per plot	65.5	1697.5	1652	409.81 ± 34.41

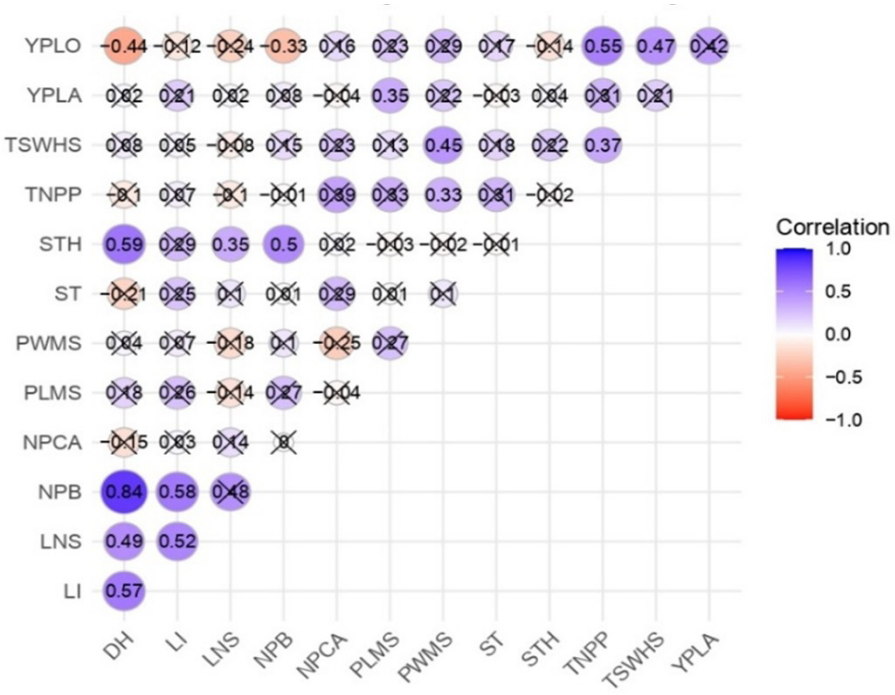


Figure 2. Pearson correlation matrix among the quantitative agromorphological variables evaluated in tarwi accessions. The colour intensity and size of the circles represent the strength and direction of the correlation: blue for positive correlations and red for negative ones. Only statistically significant correlations ($p < 0.05$) are displayed with coloured circles, while non-significant correlations are marked with an 'X'. Trait abbreviations are as in Table 3.

It was also observed that yield per plot (YPLO) shows a negative relationship with several individual traits, such as days to harvest (DH, r = -0.44). Additionally, adverse factors such as water or heat stress may further exacerbate these losses by affecting the phenological development of the plants. In this context, mechanisms such as phenological escape, the ability of some plants to complete their life cycle before unfavourable climatic conditions occur, can play a key adaptive role by indirectly reducing the number of days to harvest (Blum, 2011).

Principal component analysis (PCA)

The PCA shown in Figure 3 allowed for the synthesis of the multivariate variability present among the accessions, facilitating the identification of the most relevant traits in terms of phenotypic diversity.

In the contribution of variables to PC1 (Figure 3c), the most influential traits were days to harvest (DH), number of primary branches (NPB) and inflorescence length (LI), followed by number of leaflets per leaf (LNS) and seed thickness (STH). These variables are mainly associated with morphological and vegetative development aspects, indicating that PC1 reflects differences in plant growth and architecture. These traits are critical for adaptation to diverse agroecological conditions, as documented by Lizarazo *et al* (2010) in studies of *Lupinus* species adapted to temperate and Andean regions.

Meanwhile, PC2 was strongly influenced by yield-related variables, such as total number of pods per plant (TNPP), yield per plot (YPLO), 100-seed weight (TSWHS) and pod width on the main stem (PWMS). This dimension appears to group variables directly related to productivity, which is of great interest for breeding programmes. According to Chalampuente *et al* (2023), these traits are highly selectable and directly related to yield potential under different agroclimatic conditions. The inclusion of the variables YPLA (yield per plant) and PLMS (pod length) in this dimension further reinforces the predictive value of these variables as complementary indicators of reproductive performance, as also suggested by Galloni *et al* (2007) in yield studies of Mediterranean legumes.

Figure 4 presents a biplot derived from the PCA, summarizing the multivariate variation of the accessions based on the first two principal components (PC1 and PC2), which together explain 47.3% of the total variability (25.5% and 21.8%, respectively). This representation allows for the simultaneous visualization of the relative distribution of accessions and the influence of quantitative variables on that distribution (Jollife & Cadima, 2016).

The analysis reveals three well-defined patterns, each with distinct phenotypic and agronomic characteristics. The first pattern, located mainly in quadrant III (negative values on both dimensions), includes accessions such as PER1006570, PER1006571, PER1006006, and PER1006833, which are associated with low values for most variables, especially those with high factor loadings such as DH (days to harvest), NPB (number of primary branches) and LI (inflorescence length).

The second pattern, found in quadrant I (positive PC1 and positive PC2), groups accessions such as PER1010662, PER1010659, PER1010664, PER1010671, PER1010656, PER1006934 and PER1010661. These are associated with variables strongly loaded on PC1, such as DH, NPB, LI and STH (seed thickness), indicating a prominent morpho-structural profile and relevant agronomic potential, especially in contexts where these traits provide adaptive or productive advantages (Mohammadi & Prasanna, 2003).

The third pattern, in quadrant II (negative PC1 and

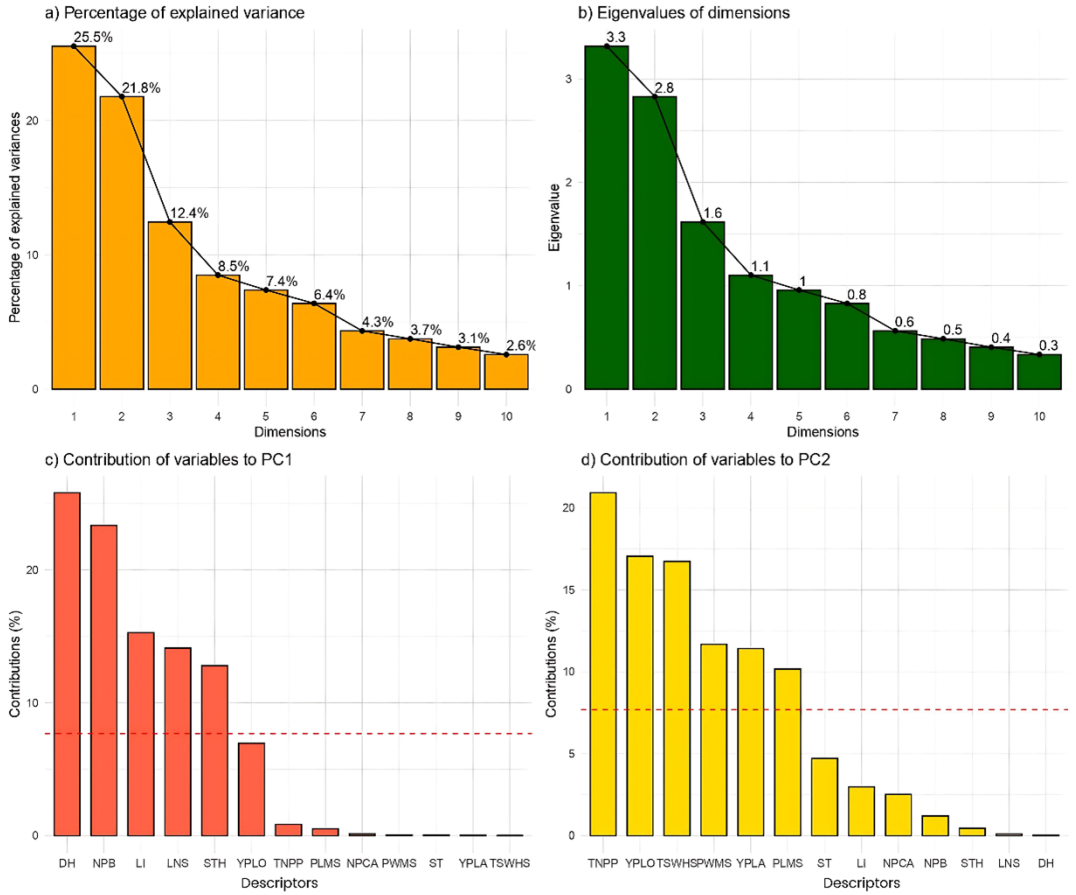


Figure 3. (a) Proportion of variance (%) of top ten principal components (PCs) , (b) eigenvalues of top 10 PCs, (c) contribution of variables to PC1 (%), and (d) contribution of variables to PC2 (%) derived from principal component analysis (PCA). Red dashed lines across bar plots are the reference lines, and the variable bars above the reference lines are considered important in contributing to the respected PCs. Trait abbreviations are as in Table 3.

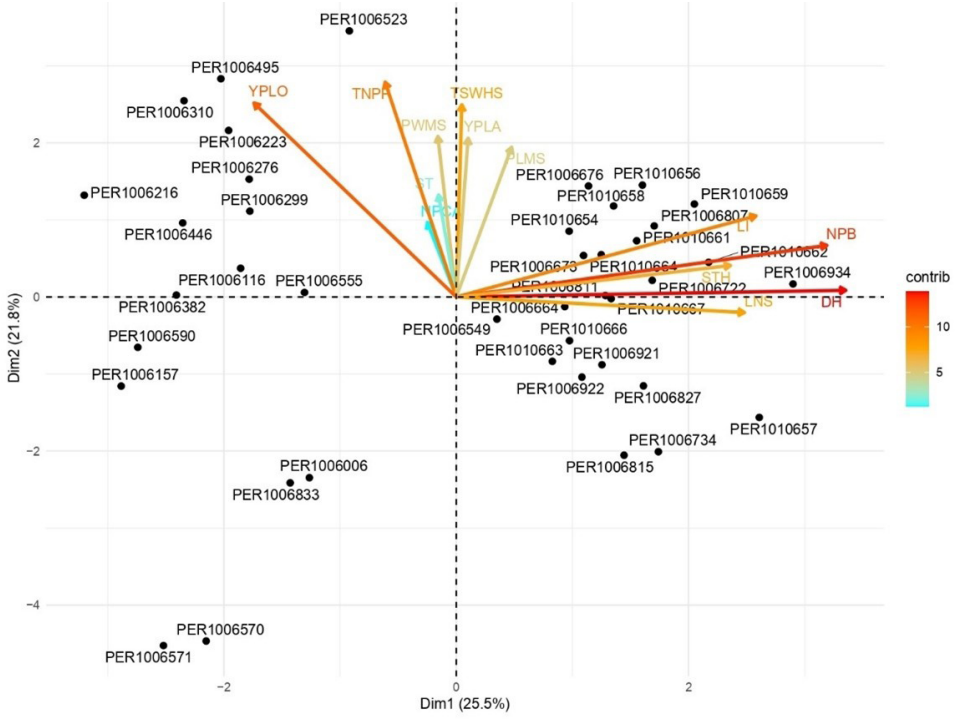


Figure 4. Principal component analysis (PCA) biplot of quantitative traits, illustrating the relative distribution of accessions (black dots) and the contribution of each trait (coloured vectors) to the first two principal components: Dim1 (25.5%) and Dim2 (21.8%), which together explain 47.3% of the total variation. The direction and length of each vector indicate the strength and influence of that trait in shaping the multivariate space, while the colour gradient from blue (low contribution) to red (high contribution) represents the relative importance of each variable in the PCA. Trait abbreviations are as in Table 3.

positive PC2), includes accessions such as PER1006495, PER1006523, PER1006310, PER1006216 and PER1006299, which are related to yield-related variables such as YPLO (yield per plot), TNPP (total number of pods per plant), TSWHS (100-seed weight) and PWMS (pod width). This profile suggests high productive potential, valuable for yield-oriented selection programmes (Bustos-Korts *et al* 2019)

Variables such as TSWHS, TNPP, LI, NPB, DH and YPLO stand out for their strong contribution to the total variability, positioning them as key factors in structuring the multivariate space. Moreover, the proximity and shared direction of vectors such as PLMS (pod length) and YPLA (yield per plant) suggest collinearity, possibly indicative of functional relationships or redundance.

Cluster dendrogram of hierarchical clustering

Figure 5 presents a dendrogram generated through hierarchical cluster analysis (HCA) using Euclidean distance and the complete linkage method, from which three phenotypically distinct clusters were identified. The first group (green) comprises the majority of the accessions, which show smaller hierarchical distances among themselves, suggesting high phenotypic similarity and possibly lower genetic divergence.

The second group (blue) includes accessions such as PER1006157, PER1006570, and PER1006590, which are characterized by a distinct configuration, possibly influenced by extreme values in certain agronomic variables.

The third group (red) comprises accessions such as PER1006495, PER1006523 and PER1006310, which show greater hierarchical distance from the other groups, indicating significant divergence in the evaluated traits. This clustering pattern supports the structure observed in PCA, reinforcing the existence of subgroups with distinct agronomic profiles.

The combined use of PCA and dendrogram analysis has been widely validated in genetic diversity and germplasm characterization studies of various agricultural species, such as common bean (Özkan *et al*, 2022), garlic (Pasupula *et al*, 2024) and rice (Nascimento *et al*, 2011), proving useful for optimizing parent selection in breeding programmes. Moreover, this hierarchical organization provides a basis for both *in situ* and *ex situ* conservation planning and the identification of promising genotypes for breeding, as highlighted in quinoa studies by Delgado *et al* (2024).

Heat map clustering

Figure 6 presents a hierarchical clustering analysis classifying the 41 accessions into three main clusters based on standardized trait values. Cluster 1 (blue bar) comprises accessions such as PER1006934, PER1006862, PER1006722 and PER1010664, characterized by intermediate to high values across structural and yield-related traits. Cluster 2 (pink bar) shows greater phenotypic heterogeneity, with elevated DH and PLMS, while Cluster 3 includes accessions PER1006310, PER1006495, PER1006299 and PER1006523, distinguished by high TNPP and TSWHS.

The heatmap reveals cluster-dependent trait associations. While some accessions show coordinated high expression of TNPP, YPLA and YPLO, consistent with productivity models in grain legumes (Clements & Cowling, 1994), this pattern is

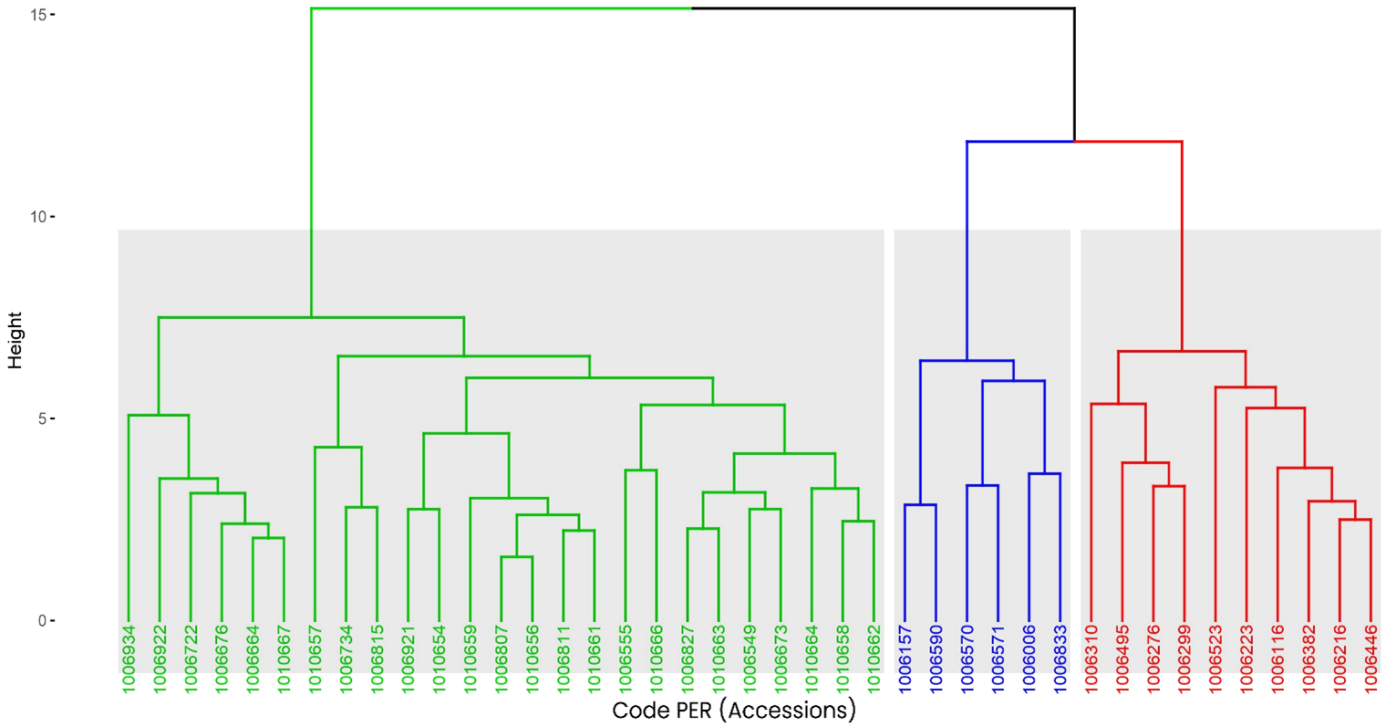


Figure 5. Dendrogram derived from Hierarchical Cluster Analysis of tarwi accessions, based on Euclidean distance and the complete linkage method. The analysis reveals three main clusters, represented by green, blue, and red branches. Each group reflects varying levels of phenotypic similarity, with shorter branch lengths indicating greater similarity among accessions.

not universal. Several genotypes exhibit high TNPP without proportionally elevated yield, indicating that pod number alone does not determine productivity. This decoupling likely reflects compensatory effects related to seed weight, pod fill efficiency, or seed abortion rates.

Morphological traits (LI, LNS) exhibit independent patterns from yield variables, suggesting that vegetative vigour and reproductive output are not strongly linked in this germplasm, consistent with observations in other legumes (Mousavi-Derazmahalleh *et al*, 2018). DH displays considerable variation with no clear association with productivity traits. Notably, several early-maturing accessions demonstrate competitive yields, indicating that earliness and productivity are not mutually exclusive, with important implications for breeding programmes in high-Andean regions with short growing seasons (Tohme *et al*, 1995)

Figure 6 complements the global correlation analysis (Figure 2) by revealing cluster-specific trait relationships, providing valuable insights for targeted germplasm selection based on specific breeding objectives.

Group description

The comparison among the three clusters (Table 5) reveals clearly differentiated phenotypic profiles among the evaluated tarwi accessions. Cluster 2 includes the most productive accessions, with high values for yield per plant (78.73g), yield per plot (746.4kg/ha), and number of pods per plant (82.01), positioning them as promising candidates for yield-oriented breeding programmes. Similar clustering approaches have been successfully applied in other crops, such as cucumber (*Cucumis sativus*), to identify high-yielding genotypes (Serhiienko *et al*, 2023).

Cluster 3, in contrast, includes accessions with enhanced vegetative growth, characterized by a longer crop cycle (269.68 days), taller plants, and a greater number of branches and pods. Although their yields are moderate, these traits offer advantages in sustainable agricultural systems, especially in the Andean highlands, where long-cycle crops help maintain soil cover, increase organic matter inputs, and

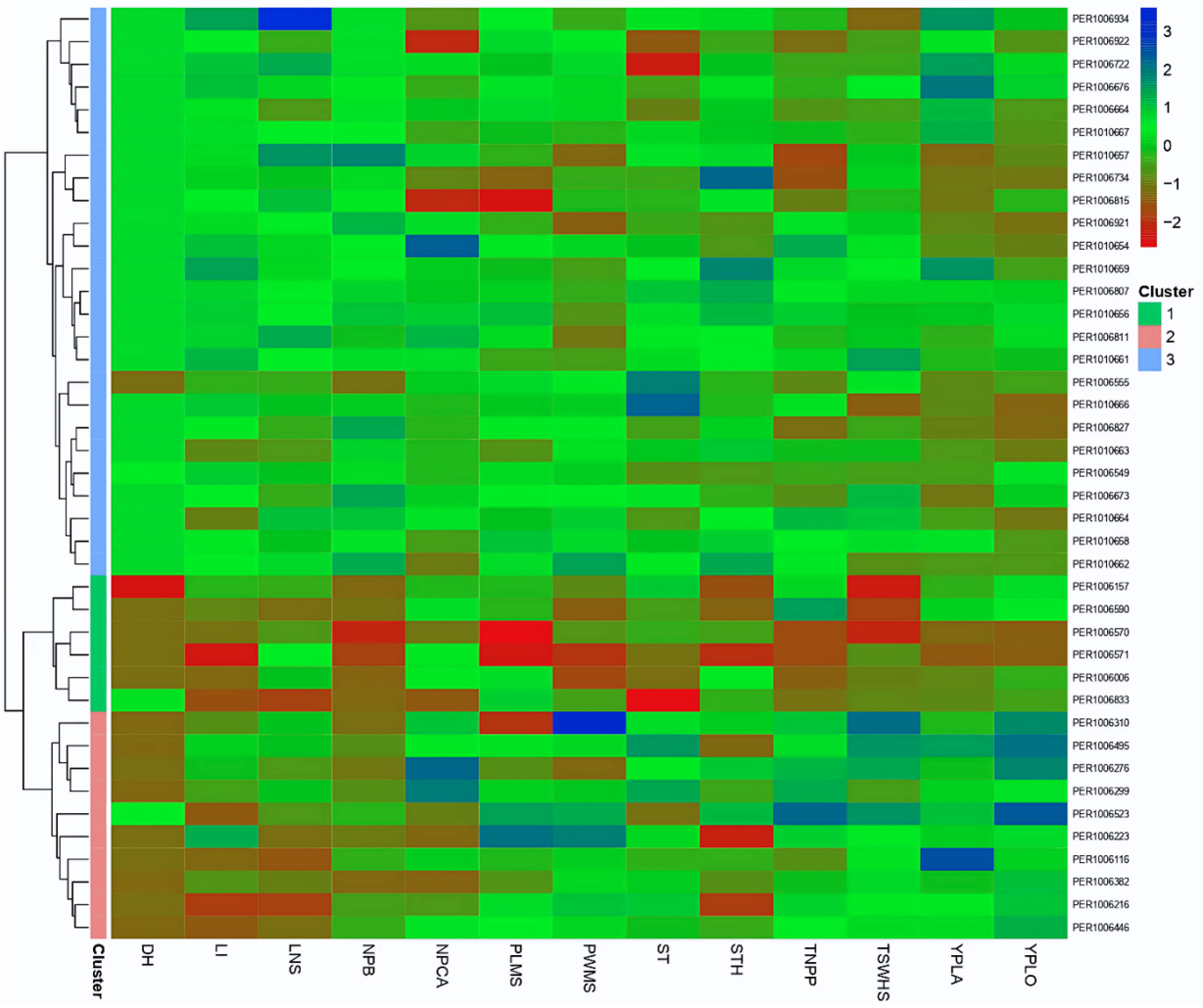


Figure 6. Heatmap with hierarchical clustering of tarwi accessions based on quantitative traits. The dendrogram classifies 41 accessions into three main clusters (coloured bars: blue, Cluster 1; pink, Cluster 2; green, Cluster 3). The heatmap displays standardized values (z-scores) of 14 morphological and agronomic variables. The colour scale ranges from blue (low values) through green (intermediate) to red (high values), revealing cluster-specific trait associations.

promote biological nitrogen fixation. These features make them particularly valuable in crop rotation schemes or in degraded soils (Jacobsen & Mujica, 2006).These accessions exhibit a more robust plant architecture, characterized by tall stature, greater number of primary branches, and an extended vegetative period, consistent with findings reported by Camarena (2012).

Cluster 1 includes early maturing accessions (243.58 days) with smaller plant size and lower yield. Nonetheless, these accessions could be strategically important in environments affected by abiotic stress, functioning as escape-type varieties that complete their life cycle before the onset of critical stress periods, a strategy also observed in cereals such as wheat and barley (Tambussi, 2006).

Altogether, these results highlight the functional diversity of the studied materials and provide valuable information for selecting accessions according to different production goals.

Regarding the comparison of the three clusters using analysis of variance, we found that the majority of morphological descriptors exhibited clear differences among groups. Of the 13 traits evaluated, eleven were statistically different among clusters, while only two remained relatively uniform. This suggests that the three groups represent truly morphologically distinct entities, each with its own characteristic profile.

The most determinant descriptors, such as NPB, exhibited the most pronounced differences among groups, with a progressive and consistent increase: Cluster 1 with 10.63, Cluster 2 with 12.97, and Cluster 3 with 17.51 ($p < 0.001$). This pattern demonstrates clear differentiation in plant architecture, possibly reflecting distinct adaptation strategies.

Descriptor DH also clearly differentiated the groups. Plants in Cluster 3 required significantly more time to reach maturity, approximately 270 days, compared to Clusters 1 and 2, which reached maturity in 243–245 days ($p < 0.001$). LI exhibited a similar pattern, progressively increasing from Cluster 1 (32.17) to Cluster 3 (38.81), indicating changes in foliar morphology ($p < 0.001$).

In terms of yield components, TSWHS was higher in Cluster 2 (30.87) compared to the other groups ($p < 0.001$), suggesting superior reproductive quality in this cluster. YPLO showed greater variability, with Cluster 2 being significantly superior (746.4) to Clusters 1 and 3 ($p < 0.001$). YPLA also differentiated the groups, with Cluster 2 showing considerably higher values ($p < 0.001$).

Interestingly, NPCA and ST showed no significant variation among clusters ($p > 0.05$), indicating that these traits remain relatively stable in the population regardless of cluster assignment, suggesting that these characteristics are less influenced by the factors that determine cluster differentiation.

Table 5. Cluster means and ANOVA for different descriptors. Cluster means based on the grouping generated in Figure 5. ***, $p < 0.001$; **, $p < 0.01$; *, $p < 0.05$; ns, not significant. DH, days to harvest; LI, inflorescence length; LNS, number of leaflets per leaf; NPB, number of primary branches; NPCA, number of pods on the central axis; PLMS, pod length; PWMS, pod width; ST, stem thickness; STH, seed thickness; TNPP, total number of pods per plant; TSWHS, total seed mass in hundred seeds; YPLA, yield per plant; YPLO, yield per plot.

Descriptor	Cluster 1 (n = 6)	Cluster 2 (n = 10)	Cluster 3 (n = 25)	F-value	p-value	Significance
DH	243.58 ± 12.38	244.85 ± 7.97	269.68 ± 5.65	46.594	0.0000	***
LI	32.17 ± 2.53	34.35 ± 3.47	38.81 ± 2.21	27.631	0.0000	***
LNS	8.22 ± 0.29	8.14 ± 0.22	8.58 ± 0.32	3.333	0.0463	*
NPB	10.63 ± 1.26	12.97 ± 1.24	17.51 ± 1.88	66.852	0.0000	***
NPCA	15.91 ± 1.44	16.65 ± 2.16	16.18 ± 1.6	0.823	0.4464	ns
PLMS	9.18 ± 0.7	9.55 ± 0.53	9.52 ± 0.35	4.246	0.0216	*
PWMS	1.49 ± 0.05	1.65 ± 0.11	1.59 ± 0.06	5.181	0.0102	*
ST	6.68 ± 0.59	7.28 ± 0.4	7.12 ± 0.47	1.924	0.1598	ns
STH	5.07 ± 0.29	5.18 ± 0.32	5.45 ± 0.23	3.420	0.0430	*
TNPP	57.08 ± 21.8	82.01 ± 13.89	65.95 ± 14.4	7.777	0.0014	**
TSWHS	24.66 ± 1.83	30.87 ± 2.19	28.6 ± 1.85	22.310	0.0000	***
YPLA	36.02 ± 19.52	78.73 ± 28.99	59.3 ± 33.35	11.150	0.0001	***
YPLO	334.83 ± 222.45	746.4 ± 175.05	325.97 ± 154.76	23.747	0.0000	***

Qualitative variables of the tarwi collection

Phenotypic characteristics of flowers and seeds of representative accessions grouped by clusters

Figure 7 illustrates the relationship between the evaluated phenotypic traits, highlighting that Cluster 1 was composed exclusively of accessions with white seeds and blue flowers. Among these, the most representative were PER1006833 and PER1006006, indicating high uniformity in these descriptors.

Similarly, Cluster 2 comprised only accessions with white seeds. However, although most (n = 9) had blue flowers, one particular accession (PER1006310) exhibited pink flowers, revealing a slight variation in this floral trait.

In contrast, Cluster 3 was characterized by greater variability in seed colour, including accessions with white, black, brown, and dark-toned seeds. Of the 25 accessions in this cluster, 8 showed seed colours other than white, while 17 retained the white colour. Despite this seed colour diversity, all accessions in Cluster 3 shared the same flower colour, blue.

Frequency analysis and Shannon index

The qualitative traits of tarwi are presented in Table 6, where 14 descriptors were evaluated. Three traits (SL, CCSSJOF and IRCSJOF) were excluded due to being monomorphic and not contributing to morphological variability. The results included the conversion of phenotypic classes and their estimated diversity using the Shannon index (H′), revealing significant variability among the analyzed traits, particularly in flower colour, pubescence of mature pods, and seed shape.

For petiole pigmentation (PGPE), 92.68% of the accessions showed pigment presence. Regarding the intensity of the flower bud colour before opening (IFBCJBO), the ‘medium’ class was predominant (51.22%), followed by ‘dark’ (21.95%), ‘light’ (14.63%), and ‘very dark’ (12.20%). This distribution corresponds to the colour of the flower wing before opening (FWCJBO), which was blue in 97.56% of the accessions and pink in only one (2.44%).

The flower keel before opening (IFKCJBO) predominantly exhibited pale tones, with 70.73% of the ‘very light’ category. Upon wilting, the colour of the flower wing (FWCJBW), the marginal band of the standard petal (MBCSFJBW), and the flower keel (FKCJBW) remained blue or purple in almost all accessions, while the colour of the central spot of the standard (CCSSFJBW) was more diverse: purple (82.93%), lilac (14.63%) and white (2.44%).

The colour of the intermediate region of the standard petal after wilting (IRCSFJBW) was mostly orange (56.10%), followed by brown (39.02%) and yellow (4.88%). Flowering uniformity (UFP) was high, with 68.29% of accessions

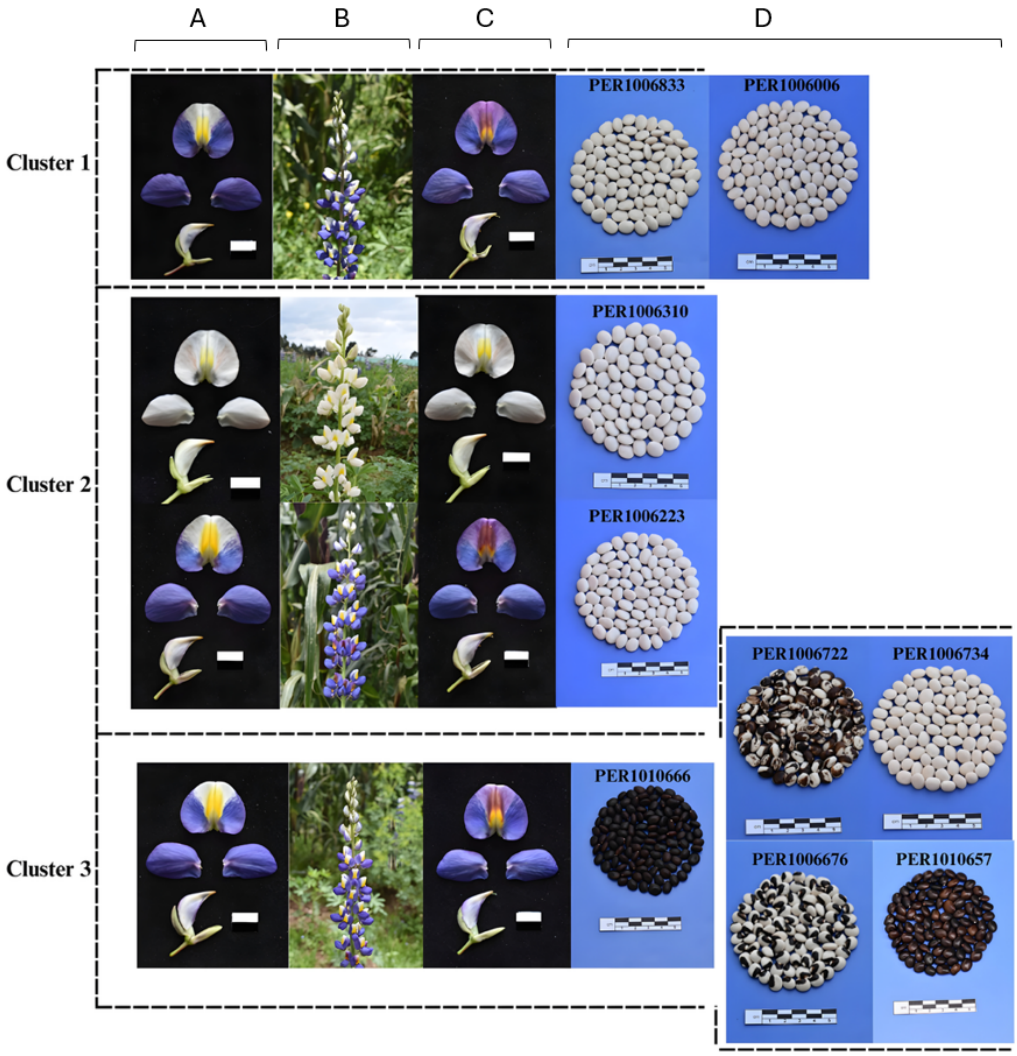


Figure 7. Variability of the accessions according to cluster grouping. (A) Dissected flower before wilting; (B) central inflorescence in the field; (C) dissected flower after wilting; (D) seed variability present in the groups.

classified as 'very uniform'.

Regarding pubescence in mature pods (MPP), moderate pubescence was most frequent (68.29%), while the rest showed either light (12.20%) or strong pubescence (19.51%). Seed shape (SSH) was mostly oval (92.68%), with some lenticular or flattened oval variants. The primary seed colour (PSC) was predominantly white (80.49%), while the remaining accessions exhibited black (14.63%) and brown (4.88%) seeds. For seed colour intensity (IPSC), the 'medium' class predominated (80.49%).

H′ was used to quantify phenotypic variability in the descriptors. High diversity (H′ > 0.60) was observed in variables such as IFBCJBO, IRCSFJBW, MPP, IFKCJBO, UFP and PSC, indicating good potential for genetic improvement programmes. Intermediate diversity (0.40 < H′ ≤ 0.60) was found in CCSSFJBW and IPSC, while low diversity (0.10 ≤ H′ ≤ 0.40) was recorded in SSH, PGPE and several floral descriptors (FWCJBO, FWCJBW, MBCSFJBW, FKCJBW), suggesting uniformity in these traits similar to the findings of Azam *et al* (2024) in pea genotypes.

Table 6. Descriptor states, and frequency of qualitative traits, Shannon Weaver diversity index (H′), descriptor states and frequency of qualitative traits. *, A total of 14 descriptors were evaluated, as three were excluded for being monomorphic and not contributing to genetic variability.

Traits	Morphological descriptor	Rank	Phenotypic class	Frequency	H′
PGPE	Pigmentation of petioles	0	Absent	3	0.2618
		1	Present	38	
IFBCJBO	Intensity of flower bud colour just before opening	3	Light	6	1.2134
		5	Medium	21	
		7	Dark	9	
		9	Very dark	5	
FWCJBO	Flower wing colour just before opening	4	Pink	1	0.1147
		7	Blue	40	
IFKCJBO	Intensity of flower keel colour just before opening	1	Very light	29	0.7908
		3	Light	4	
		5	Medium	8	
FWCJBW	Flower wing colour just before wilting	4	Pink	1	0.1147
		7	Blue	40	
MBCSFJBW	Marginal band colour of standard of flower just before wilting	4	Pink	1	0.1147
		7	Blue	40	
CCSSFJBW	Colour of central spots of standard of flower just before wilting	1	White	1	0.5271
		8	Purple	34	
		10	Lilac	6	
IRCSFJBW	Intermediate region colour of standard of flower just before wilting	2	Yellow	2	0.8388
		3	Orange	23	
		9	Brown	16	
FKCJBW	Flower keel colour just before wilting	1	White	1	0.1147
		8	Purple	40	
UFP	Uniformity of flowering of the plot	5	Medium	13	0.6246
		7	Very much	28	
MPP	Mature pod pubescence	3	Slight	5	0.8359
		5	Medium	28	
		7	Strong	8	
SSH	Seed shape	2	Flattened spherical or lenticular	2	0.3083
		3	Oval	38	
		4	Flattened oval	1	
PSC	Primary seed colour	1	White	33	0.6033
		9	Brown	2	
		12	Black	6	
IPSC	Intensity primary seed colour	5	Medium	33	0.4936
		9	Very dark	8	

Conclusions

Based on the exploratory analysis, notable agromorphological diversity was identified, expressed in both vegetative and productive traits. Potential associations were observed between structural and reproductive variables such as number of branches, number of pods, seed thickness, and yield per plant, which may be relevant for the selection of superior materials. In particular, the negative correlation between days to harvest and yield per plot in some accessions opens the possibility of identifying early maturing accessions with good performance, which would be favorable in scenarios of climatic risk or shorter cropping seasons.

The qualitative analysis also revealed appreciable diversity in floral colours, seed shapes, and degrees of pod pubescence traits that, beyond their genetic value, could influence the cultural and commercial acceptance of the crop in different regions. Furthermore, the accessions were grouped into three distinct phenotypic clusters: one with high productive potential, another with predominance of vegetative characteristics, and a third composed of short-cycle materials with intermediate yield.

In this sense, the preliminary results of this study suggest that tarwi represents a valuable resource to guide future conservation strategies, identification of promising accessions, and the development of genetic improvement programmes. While the phenotypic groupings identified are promising, it is still necessary to expand the number of evaluated accessions and carry out complementary analyses to validate and consolidate the observed patterns. Within this framework, this study lays the foundation for more robust research efforts that will contribute to strengthening recommendations for the selection and utilization of materials with desirable agronomic profiles across diverse agroecological contexts.

Data availability statement

The full dataset in this study can be found at the following link: https://doi.org/10.5281/zenodo.15740283. Additional information can be requested from the corresponding author.

Author contributions

Kevin Ortega: Conceptualization, data curation, analyses conduction, investigation, methodology, resources, provision of study materials, validation, visualization, preparation of figures and tables, writing (original draft, review and editing); Eunice Peña: supervision, data curation, resources, provision of study materials, validation, preparation of figures and tables, writing (original draft, review and editing); Carolina Girón: resources, validation, data curation, provision of study materials; Nery Amaro: resources, validation, provision of study materials; Claudia Rios: analysis conduction, methodology, writing (review and editing); Bertha Lopez: resources, validation, provision of study materials; Francis Cerrón: analysis conduction, methodology; Steve Camargo: supervision, conceptualization, visualization, writing (review and editing), acquisition of financial support; Samuel Pizarro: supervision, conceptualization, provision of study materials, formal analysis, investigation, resources, validation.

All authors read and approved the final manuscript.

Conflict of interest statement

The authors confirmed that no conflict of interest exists.

Acknowledgments

The authors express their gratitude to the team actively working on the investment project of the National Institute of Agrarian Innovation 'PROAGROBIO' (CUI N° 2480490) and the Sub-Directorate of Genetic Resources. Special thanks to the Continental University, Huancayo Campus, for their constant support of the research work and to the collaborators of the Estación Experimental Santa Ana for their valuable contributions throughout the different phases of the research execution.

References

Aguilar Angulo, L. (2015). *Evaluación del rendimiento de grano y capacidad simbiótica de once accesiones de tarwi (Lupinus mutabilis S.), bajo condiciones de Otuzco - La Libertad [Undergraduate thesis, Universidad Nacional Agraria La Molina]*. https://hdl.handle.net/20.500.12996/1626.

Azam, M. G., Sarker, U., Hossain, M. A., Mahabubul Alam, A. K. M., Islam, M. R., Hossain, N., & Alamri, S. (2024). Phenotypic diversity in qualitative and quantitative traits for selection of high yield potential field pea genotypes. *Scientific Reports*, *14*(1), 1–24. https://doi.org/10.1038/s41598-024-69448-7.

Blum, A. (2011). Plant breeding for water-limited environments. *Plant Breeding for Water-Limited Environments*, 1–255. https://doi.org/10.1007/978-1-4419-7491-4.

Bustos-Korts, D., Boer, M. P., Malosetti, M., Chapman, S., Chenu, K., Zheng, B., & van Eeuwijk, F. A. (2019). Combining crop growth modeling and statistical genetic modeling to evaluate phenotyping strategies. *Frontiers in Plant Science*, *10*. https://doi.org/10.3389/FPLS.2019.01491.

Caligari, P. D. S., Römer, P., Rahim, M. A., Huyghe, C., Neves-Martins, J., & Sawicka-Sienkiewicz, E. J. (2000). *The potential of Lupinus mutabilis as a crop*. In Knight, R. (eds) Linking Research and Marketing Opportunities for Pulses in the 21st Century (pp. 569–573). Springer, Dordrecht. https://doi.org/10.1007/978-94-011-4385-1_54.

Camarena Mayta, F., Huaringa Joaquín, A., Jiménez Dávalos, J., & Mostacero Neyra, E. (2012). Revaloración de un cultivo subutilizado: chocho o tarwi (*Lupinus mutabilis* S.) (Vol. 1). CONCYTEC. https://hdl.handle.net/20.500.12390/3789.

Cano, A. F. S., Dávalos, J. J., Bonilla, C. F. L., Espinoza, E. R., & Oblitas, Y. F. A. (2022). Caracterización intragenotípica de siete accesiones de tarwi (*Lupinus mutabilis* S.) usando marcadores moleculares ISSR. Hatun Yachay Wasi, 1(2), 69–82. https://doi.org/10.57107/HYW.V1I2.25.

Chalampuente, D., Tapia, C., & Sorensen, M. (2021). The Andean Lupine-'El Chocho' or 'Tarwi' (*Lupinus mutabilis* S.). *Biodiversity Online Journal*, *1*(4). https://doi.org/10.31031/BOJ.2021.01.000520.

Chalampuente-Flores, D., Mosquera-Losada, M. R., Ron, A. M. D., Tapia Bastidas, C., & Sørensen, M. (2023). Morphological and ecogeographical diversity of the Andean lupine (*Lupinus mutabilis* S.) in the high Andean region of Ecuador. *Agronomy, 13*(8), 2064. https://doi.org/10.3390/agronomy13082064.

Clements, J. C., & Cowling, W. A. (1994). Patterns of morphological diversity in relation to geographical origins of wild *Lupinus angustifolius* from the Aegean region.

Genetic Resources and Crop Evolution, *41*(2), 109–122. https://doi.org/10.1007/BF00053055.

Delgado, H., Tapia, C., Manjarres-Hernández, E. H., Borja, E., Naranjo, E., & Martín, J. P. (2024). Phenotypic diversity of quinoa landraces cultivated in the Ecuadorian Andean region: In situ conservation and detection of promising accessions for breeding programs. *Agriculture,* 14(3), 336. https://doi.org/10.3390/agriculture14030336.

Enrique Quispe, E. (2022). Evaluación de las propiedades fisicoquímicas de la harina de Lupinus mutabilis S. (tarwi) desamargado [Undergraduate thesis, Universidad Nacional de Huancavelica]. https://hdl.handle.net/20.500.14597/4826.

Enriquez, L., Ortega, K., Ccopi, D., Rios, C., Urquizo, J., Patricio, S., Alejandro, L., Oliva-Cruz, M., Barboza, E., & Pizarro, S. (2025). Detecting changes in soil fertility properties using multispectral UAV images and machine learning in central Peru. *AgriEngineering, 7*(3), 70. https://doi.org/10.3390/agriengineering7030070.

Eticha, F., Bekele, E., Belay, G., & Börner, A. (2005). Phenotypic diversity in tetraploid wheats collected from Bale and Wello regions of Ethiopia. *Plant Genetic Resources*, *3*(1), 35–43. https://doi.org/10.1079/pgr200457.

Everitt, B. S., Landau, S., Leese, M., & Stahl, D. (2011). *Cluster Analysis* (5th ed.), 1–330. https://doi.org/10.1002/9780470977811.

Food and Agriculture Organization of the United Nations (FAO). (2010). The contribution of plant genetic resources for food and agriculture to food security and sustainable agricultural development. *The Second Report on the State of the World's Plant Genetic Resources for Food and Agriculture* (pp. 182–201). https://www.fao.org/4/i1500e/i1500e08.pdf

Food and Agriculture Organization of the United Nations (FAO). (2019). *The state of the world's biodiversity for food and agriculture* (J. Bélanger & D. Pilling, Eds.). FAO Commission on Genetic Resources for Food and Agriculture Assessments. http://www.fao.org/3/CA3129EN/CA3129EN.pdf.

Flores, H. E., Walker, T. S., Guimarães, R. L., Bais, H. P., & Vivanco, J. M. (2003). Andean root and tuber crops: Underground rainbows. *HortScience*, *38*(2), 161–167. https://doi.org/10.21273/HORTSCI.38.2.161.

Galloni, M., Podda, L., Vivarelli, D., & Cristofolini, G. (2007). Pollen presentation, pollen-ovule ratios, and other reproductive traits in Mediterranean Legumes (Fam. Fabaceae - Subfam. Faboideae). *Plant Systematics and Evolution*, *266*(3), 147–164. https://doi.org/10.1007/S00606-007-0526-1.

Gresta, F., Wink, M., Prins, U., Abberton, M., Capraro, J., Scarafoni, A., & Hill, G. (2017). Lupins in European cropping systems. *Legumes in Cropping Systems*, 88–108. https://doi.org/10.1079/9781780644981.0088.

Guilengue, N., Alves, S., Talhinhas, P., & Neves-Martins, J. (2019). Genetic and genomic diversity in a tarwi (*Lupinus mutabilis* S.) germplasm collection and adaptability to Mediterranean climate conditions. *Agronomy*, *10*(1), 21. https://doi.org/10.3390/agronomy10010021.

Gulisano, A., Alves, S., Martins, J. N., & Trindade, L. M. (2019). Genetics and breeding of *Lupinus mutabilis*: An emerging protein crop. *Frontiers in Plant Science, 10,* 1385. https://doi.org/10.3389/fpls.2019.01385.

Gulisano, A., Alves, S., Rodriguez, D., Murillo, A., van Dinter, B. J., Torres, A. F., Gordillo-Romero, M., Torres, M. de L., Neves-Martins, J., Paulo, M. J., & Trindade, L. M. (2022). Diversity and agronomic performance of *Lupinus mutabilis* germplasm in European and Andean environments. *Frontiers in Plant Science*, *13*, 903661. https://doi.org/10.3389/fpls.2022.903661.

International Board for Plant Genetic Resources (IBPGR). (1981). Lupin descriptors (Vol. 1). https://hdl.handle.net/10568/73417.

Jacobsen, S.-E., & Mujica, A. (2006). El tarwi (Lupinus mutabilis S.) y sus parientes silvestres. In M. Moraes R., B. Øllgaard, L. P. Kvist, F. Borchsenius, & H. Balslev (Eds.), Botánica Económica de Los Andes Centrales, (pp. 458–482). Universidad Mayor de San Andrés.

Jollife, I. T., & Cadima, J. (2016). Principal component analysis: A review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *374*(2065). https://doi.org/10.1098/RSTA.2015.0202.

Kassambara, A., & Mundt, F. (2020). *factoextra: Extract and visualize the results of multivariate data analyses* [version 1.0.7]. *CRAN: Contributed Packages*. https://doi.org/10.32614/CRAN.PACKAGE.FACTOEXTRA.

Lê, S., Josse, J., & Husson, F. (2008). FactoMineR: An R package for multivariate analysis. *Journal of Statistical Software*, *25*(1), 1–18. https://doi.org/10.18637/JSS.V025.I01.

Lizarazo, C., Stoddard, F., Mäkelä, P., & Santanen, A. (2010). Genetic variability in the physiological responses of Andean lupin to drought stress. *Suomen Maataloustieteellisen Seuran Tiedote*, *26*, 1–5. https://doi.org/10.33354/smst.76862.

Massawe, F. J., Mayes, S., Cheng, A., Chai, H. H., Cleasby, P., Symonds, R., Ho, W. K., Siise, A., Wong, Q. N., Kendabie, P., Yanusa, Y., Jamalluddin, N., Singh, A., Azman, R., & Azam-Ali, S. N. (2015). The Potential for underutilised crops to improve food security in the face of climate change. *Procedia Environmental Sciences,* 29, 140–141. https://doi.org/10.1016/j.proenv.2015.07.228.

Mohammadi, S. A., & Prasanna, B. M. (2003). Analysis of genetic diversity in crop plants—salient statistical tools and considerations. *Crop Science,* 43(4), 1235–1248. https://doi.org/10.2135/cropsci2003.1235.

Morales-Casco, L. A., & Zúniga-González, C. A. (2016). Impactos del cambio climático en la agricultura y seguridad alimentaria. *Revista Iberoamericana de Bioeconomía y Cambio Climático*, 2(1), 269–291. https://doi.org/10.5377/ribcc.v2i1.5700.

Mousavi-Derazmahalleh, M., Bayer, P. E., Nevado, B., Hurgobin, B., Filatov, D., Kilian, A., Kamphuis, L. G., Singh, K. B., Berger, J. D., Hane, J. K., Edwards, D., Erskine, W., & Nelson, M. N. (2018). Exploring the genetic and adaptive diversity of a pan-Mediterranean crop wild relative: narrow-leafed lupin. *Theoretical and Applied Genetics*, *131*(4), 887–901. https://doi.org/10.1007/S00122-017-3045-7.

Mujica, Á., Chura, E., Moscoso, G., Chuquimia, D., Romero, T., Astete, A., Calandri, E., & Montoya, P. (2021). Selección de cultivares de tarwi (*Lupinus mutabilis* S.) por rendimiento, precocidad, contenido de aceite y proteína en Puno, Perú. *Agrárias: Pesquisa e Inovação Nas Ciências Que Alimentam o Mundo* VI (pp. 1–13). https://doi.org/10.37572/EDART_3004213541.

Nascimento, W. F. Do, da Silva, E. F., & Veasey, E. A. (2011). Agro-morphological characterization of upland rice accessions. *Scientia Agricola, 68(6), 652–660.* https://doi.org/10.1590/S0103-90162011000600008.

Ojuederie, O. B., Igwe, D. O., Ludidi, N. N., & Ikhajiagbe, B. (2023). Editorial: Neglected and underutilized crop species for sustainable food and nutritional security: prospects and hidden potential. *Frontiers in Plant Science*, *14*, 1358220. https://doi.org/10.3389/fpls.2023.1358220.

Özkan, G., Haliloğlu, K., Türkoğlu, A., Özturk, H. I., Elkoca, E., & Poczai, P. (2022). Determining genetic diversity and population structure of common bean (*Phaseolus vulgaris* L.) landraces from Türkiye using SSR markers. *Genes*, *13*(8), 1410. https://doi.org/10.3390/genes13081410.

Padulosi, S., Heywood, V., Hunter, D., & Jarvis, A. (2011). Underutilized species and climate change: current status and outlook. In S. S. Yadav, R. J. Redden, J. L. Hatfield, H. Lotze-Campen, & A. E. Hall (Eds.), *Crop Adaptation to Climate Change* (pp. 507–521). https://doi.org/10.1002/9780470960929.CH35.

Palacios-López, L. A. (2024). Nuevas tendencias e innovaciones para una agricultura sostenible. *Horizon Nexus Journal*, *2*(2), 15–28. https://doi.org/10.70881/hnj/v2/n2/36.

Pasupula, K., Verma, P., Zimik, M., Kaur, C., Vasudev, S., & Khar, A. (2024). Morphological, biochemical and molecular characterization of short-day tropical Indian garlic (*Allium sativum* L.). *Heliyon,* 10(18), e37553. https://doi.org/10.1016/j.heliyon.2024.e37553.

Pearson, K. (1895). VII. Note on regression and inheritance in the case of two parents. *Proceedings of the Royal Society of London*, *58*(347–352), 240–242. https://doi.org/10.1098/rspl.1895.0041.

Peralta I., E., Murillo I., A., Mazón, N., Villacrés, E., & Rivera M., M. (2013). Catálogo de variedades mejoradas de granos andinos: Chocho, quinua y amaranto, para la sierra de Ecuador. http://repositorio.iniap.gob.ec/handle/41000/2713.

R Core Team (2023). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. url: https://www.R-project.org/.

Saxena, K. B. (2018). SyBrid - a new breeding method for food legumes. International Journal of Science and Research. http://oar.icrisat.org/id/eprint/11562.

Servicio Nacional de Meteorología e Hidrología bdel Perú (SENAMHI). (2024). SENAMHI - Estaciones. https://www.senamhi.gob.pe/main.php?dp=junin&p=estaciones.

Serhiienko, O. V., Solodovnyk, L. D., Harbovska, T. M., & Radchenko, L. O. (2023). Assessments of structural yield indicators of cucumber genotypes using cluster analysis. *Vegetable and Melon Growing*, *74*, 33–39. https://doi.org/10.32717/0131-0062-2023-74-33-39.

Seyedsayamdost, E. (2020). Sustainable development goals. *Essential Concepts of Global Environmental Governance* (pp. 251–253). https://doi.org/10.4324/9780367816681-102.

Tambussi, E. A. (2006). Fotosíntesis, fotoprotección, productividad y estrés abiótico: algunos casos de estudio [Doctoral dissertation, Universidad de Barcelona]. https://dialnet.unirioja.es/servlet/tesis?codigo=3773&info=resumen&idioma=SPA.

Tapia, M. (2015). El tarwi, lupino andino (1st ed.). Fondo Italo Peruano. https://fadvamerica.org/wp-content/uploads/2017/04/TARWI-espanol.pdf

Tohme, J., Toro, O. C., Vargas, J., & Debouck, D. G. (1995). Variability in Andean nuña common beans (*Phaseolus vulgaris*, Fabaceae). *Economic Botany,* 49(1), 78–95. https://doi.org/10.1007/BF02862280.

Vera-Vega, M., Jimenez-Davalos, J., & Zolla, G. (2022). The micronutrient content in underutilized crops: the *Lupinus mutabilis* S. case. *Scientific Reports*, *12*(1). https://doi.org/10.1038/S41598-022-19202-8.

Wilkinson, L., & Friendly, M. (2009). History corner the history of the cluster heat map. *American Statistician,* 63(2), 179–184. https://doi.org/10.1198/TAS.2009.0033.

Genetic Resources

ORIGINAL ARTICLE

Genetic Resources (2026), 7 (13), 57–64
DOI: 10.46265/genresj.GUWQ7085
https://www.genresj.org
ISSN: 2708-3764

A comprehensive study on how inbreeding influences the growth and reproductive traits of six indigenous chicken breeds subjected to selection programmes

Saber Jelokhani-Niaraki and Sholeh Ghorbani*

Biotechnology Research Department, Animal Science Research Institute of Iran, Agricultural Research, Education and Extension Organization (AREEO), Karaj, Iran

Abstract: Indigenous chickens are a significant element of the farming system in rural areas of Iran. This study presents a systematic analysis of how inbreeding affects the growth and reproductive traits across six indigenous chicken breeds that are under genetic selection programmes. Pedigree data of 404,597 chickens from six indigenous chicken breeding centres were collected and analyzed over 15 to 29 generations. The study included eight production and reproduction traits. The results showed that the average inbreeding coefficient in the studied populations varied between 2.2% to 6.3% in centres. The average inbreeding rate was estimated to be between 0.3% and 0.6%, which is within the acceptable range for breeding programmes. Regression analysis of studied traits on inbreeding percentage showed that increased inbreeding had a slight negative effect on some traits, such that every 1% increase in inbreeding resulted in a decrease of 1.53 to 3.51g in body weight at 12 weeks and an increase of 0.12 to 0.38 days in age at sexual maturity. However, the effect of inbreeding on egg traits was insignificant. In conclusion, despite the implementation of a closed breeding system and genetic selection in centres, inbreeding has increased slowly in the populations, and genetic diversity has been maintained at an adequate level due to the successful implementation of selection and mating programmes running in indigenous chicken breeding centres.

Keywords: Inbreeding, indigenous chickens, selection, genetic progress, Iran

Citation: Jelokhani-Niaraki, S. and Ghorbani, S. (2026) "A comprehensive study on how inbreeding influences the growth and reproductive traits of six indigenous chicken breeds subjected to selection programmes", *Genetic Resources*, 7(13), pp. 57–64. doi: 10.46265/genresj.GUWQ7085.

Introduction

Indigenous chicken (IC) breeds significantly contribute to rural economies in many developing and underdeveloped countries. They are especially important for the rural poor and marginalized communities, providing supplementary income and nutritious chicken eggs and meat for their consumption (Padhi, 2016). ICs are characterized by their disease resistance, adaptability to various climates, high immune competence and desirable meat and carcass quality (Jaturasitha *et al*, 2008; Szalay *et al*, 2016; Radwan, 2020). The production system's adaptability to diverse agroecological conditions, coupled with IC's minimal resource requirements, accounts for their widespread adoption across various regions (Milkias *et al*, 2019). The majority of local populations have evolved in response to a specific and often challenging environment (Tolone *et al*, 2023). The adaptability of ICs to different environmental conditions, along with their proven potential for breeding improvements, offers a valuable genetic resource for tackling the challenges of food security in a world affected by climate change and increasing human population (Lawal and Hanotte, 2021). Although these breeds possess

* Corresponding author: Sholeh Ghorbani (sholehghorbani@gmail.com)

Received: 08.07.2025 | Accepted: 30.12.2025 | Published online: 09.02.2026

significant characteristics, they are not commonly farmed for commercial use because their productivity is lower than that of commercial breeds (Buranawit *et al*, 2025). Nevertheless, there has been a growing consumer demand for natural and organic food products in recent years. This shift has impacted the poultry sector, resulting in a growing popularity of alternative rearing methods like organic, agroforestry, and free-range systems (Stefanetti *et al*, 2023). Nearly 90% of rural households traditionally raise a small flock of indigenous chickens in free-range, semi-scavenging systems (Padhi, 2016).

The productive performance of ICs can be enhanced through genetic improvement and optimized husbandry practices, including advancements in feeding, veterinary care and farm management protocols. From a genetic perspective, improvement strategies can involve selective breeding, crossbreeding, or a hybrid approach that integrates both methods (Padhi, 2016). ICs serve as reservoirs of genetic diversity, providing the foundation for selection in various production situations (Desta *et al*, 2013; Lawal *et al*, 2018). Thus, selective breeding aimed at improving traits in IC breeds may result in improved productivity within free-range systems, without raising production costs or losing biodiversity (Magothe *et al*, 2012). Although the selection programmes exhibit lengthy periods to achieve results, they provide more permanent genetic solutions, and such results are long-lasting (Padhi, 2016).

In Iran, backyard ICs significantly contribute to rural agriculture by serving as a source of both protein and income. Due to their importance in sustainable agriculture and consumer preference for IC products over commercial breeds, efforts have intensified to improve IC productivity. To achieve this, indigenous chicken breeding centres (ICBCs) were established across six provinces (Esfahan, Khorasan Razavi, Fars, Yazd, West Azerbaijan, and Mazandaran) in the country. The focus of ICBCs lies in the genetic improvement of some economic traits, alongside the reproduction and distribution of dual-purpose chickens that are adapted to regional conditions (for both rural and semi-industrial contexts). The IC breeding programme at these centres has been underway for several years to improve growth and reproductive traits. A similar selective breeding programme is implemented at the centres, with the selection process for each generation based on the estimated breeding values for five traits under selection derived from a multi-trait analysis approach. All populations of the ICBCs are closed, with no gene flow from external sources (Jelokhani-Niaraki and Ghorbani, 2021). The original base populations from each centre have now diverged into distinct, genetically improved breeds, characterized by their improved productive and reproductive performance. The country currently has six genetically improved chicken breeds, each originating from a dedicated centre within its respective province. These registered breeds are designated as follows: the Caspian (originating from the Mazandaran region), Parseh (from Fars), Sepahan (from Esfahan), Tusika (from Khorasan Razavi), Urmiana (from West Azerbaijan) and Isatis (from Yazd).

Increased inbreeding is one of the potential challenges that may occur in closed populations. Closed populations that are small and simultaneously under selection pressures may experience a rapid decrease in heterozygosity and allelic diversity (Selvaggi *et al*, 2010). Inbred individuals are often less adaptable to environmental changes, leaving them more sensitive and weaker (Barros *et al*, 2017). Maintaining genetic diversity in a population can be achieved by minimizing the rise in inbreeding over successive generations (D'Ambrosio *et al*, 2019). Even though inbreeding can be used as a valuable tool to identify and eliminate deleterious recessive alleles in a population, its consequences, such as reduced reproductive efficiency and growth rate, increased mortality rates, and increased incidence of hereditary diseases, have raised concerns (Yadav *et al*, 2019). Limiting inbreeding is vital to maintain genetic diversity, which in turn allows future generations to adapt to environmental change and respond to selection. Without this diversity, their adaptive capacity is critically diminished (Van Wyk *et al*, 2009). Inbreeding depression in domestic animals can reduce selection response and potential genetic progress in economic traits (Selvaggi *et al*, 2010). Increasing inbreeding within the closed populations of ICBCs might unfavourably influence the traits productivity during the selection process and potentially hinder breeding progress. Consequently, tracking inbreeding levels across generations is essential for optimizing selection and mating strategies. This study examined inbreeding rates and their effects on growth and reproductive traits in six indigenous chicken breeds to guide future breeding efforts.

Materials and methods

Data and traits

Data from 404,597 chickens were collected across six ICBCs in Esfahan, Khorasan Razavi, Fars, Mazandaran, Yazd, and West Azerbaijan provinces, spanning 15 to 29 generations (a year is equivalent to each generation). The study evaluated eight traits including body weight at hatch (BW1), body weight at 8 weeks (BW8), body weight at 12 weeks (BW12), age at sexual maturity (ASM), weight at sexual maturity (WSM), egg weight on the first day of laying (EW1), egg number (EN) and average egg weight (AEW). The AEW represents the average weight of eggs produced at 28, 30 and 32 weeks, while EN indicates the average number of eggs produced over the first 84 days. To avoid the increase of inbreeding within the ICBCs, a controlled mating system has been implemented. This system maintains a rooster-to-hen ratio of approximately 1:10 and prioritizes the selection of breeding stock with minimal genetic relatedness.

After mating the hens with a specific rooster, all eggs were individually marked with the parents' identification. Eggs from each hen were housed in distinct, partitioned sections of the hatching baskets. The use of covered baskets prevented the chicks from mixing after hatching, which guaranteed that every chick's parentage was known. The chicks were then tagged individually.

At these centres, chickens were selected in each generation based on estimated breeding values (EBVs) for some economic traits, using a multi-trait animal model analysis. A detailed description of the pedigree data investigated in this study is presented in Table 1.

Table 1. Pedigree data information used in the study

Pedigree information	Indigenous chicken breeding centres (ICBCs)					
	Fars	Khorasan Razavi	Mazandaran	Esfahan	Yazd	West Azerbaijan
No. of total chickens	65,268	39,280	82,265	98,064	57,746	61,974
Inbred chickens	40,184	21,892	59,075	77,313	45,351	54,982
Sires	2,782	1,165	2,250	1,911	1,071	1,792
Dams	14,855	8,015	16,507	11,512	8,578	9,859
Chickens with offspring	17,637	9,180	18,757	13,423	9,649	11,651
Chickens without offspring	47,631	30,100	63,508	84,641	48,097	50,323
No. of generations	25	12	26	21	15	21
Years covered	1991–2017	2006–2017	1990–2017	1995–2018	2001–2017	1994–2017

Statistical analysis

Pedigree data were processed through sequential rounds of screening and quality checks using Foxpro (version 2.6) and MS Excel (2010 version), and faulty or outlier data points were eliminated. The exclusion was based on practical and experiential grounds, specifically the removal of clearly erroneous records resulting from data entry errors (e.g. implausible trait values) rather than a statistical threshold. The number of records removed was negligible (approximately 30 per breed) compared to the total dataset.

The analysis-ready files were then prepared for subsequent analysis. Inbreeding coefficients were calculated for all chickens in the pedigree using the CFC software (Sargolzaei *et al*, 2006). The descriptive statistics of these coefficients were estimated based on the population of inbred chickens throughout the period, percentage in the total population and generation. Since considering the information of common ancestors and the complete pedigree is crucial in estimating inbreeding coefficients, the inbreeding coefficient was estimated based on all available data and kinship relationships in the pedigree. The annual rate of inbreeding change was calculated by fitting a linear regression of inbreeding on generation using SPSS software (IBM Corp, 2017).Regression coefficients of studied traits on inbreeding percentage were estimated by Wombat software (Mayer, 2007) and restricted maximum likelihood (REML) method using six different models. Inbreeding coefficient was also included as a covariate in the model. In this study, among the six statistical models considered for each trait, the final appropriate model for each of them was selected through three methods of likelihood ratio test (LRT), Akaike's information criterion (AIC) and Bayesian information criterion (BIC). In the LRT test, the model with the highest log likelihood value was selected as the base model. Then, in order to evaluate the significant difference between the models, Q was estimated using the difference in log likelihood as follows (Lewis *et al*, 2011):

$$Q = 2(\log L_B - \log L_A)$$

Here, L_A and L_B are the likelihoods of the nested (studied) and full models, respectively. The value of L_B must be larger than or equal to that of L_A because model A is a special case of model B.

In this study, Q can adopt a chi-squared (χ^2) distribution. The calculated Q value for all models was compared with the χ^2 distribution. In general, the model with the highest log likelihood is selected as the most appropriate model, but if χ^2 is significant, it is statistically superior to the other model. In the case where the difference between the models is not statistically significant ($P < 0.05$), the simplest model can be selected as the most appropriate model. In the second and third methods, the model with the minimum BIC (Schwarz, 1978) and AIC (Akaike, 1974) was selected as the most appropriate model. The AIC and Bayesian information indices were calculated as follows (Fischer *et al*, 2004*)*:

$$AIC = -2\ (Log\ L) + 2p$$
$$BIC = -2\ (Log\ L) + p^{*}Log\ (N - r(X))$$

Log L is the log likelihood, p is the number of model parameters, N is the number of records, and (r)X is the rank of the matrix X.

In this study, regression coefficients of studied traits on inbreeding percentage were estimated based on the appropriate model. The models used in this study were as follows:

1. $y = Xb + Z_1a + e$
2. $y = Xb + Z_1a + Z_3c + e$
3. $y = Xb + Z_1a + Z_2m + e$ $\quad Cov\ (a,m) = 0$
4. $y = Xb + Z_1a + Z_2m + e$ $\quad Cov\ (a,m) \neq 0$
5. $y = Xb + Z_1a + Z_2m + Z_3c + e$ $\quad Cov\ (a,m) = 0$
6. $y = Xb + Z_1a + Z_2m + Z_3c + e$ $\quad Cov\ (a,m) \neq 0$

where y: observations vector; X: incidence matrix that relate observations to the fixed effects of model; b: the vector of fixed effects and associated variables (including the fixed effects generation-hatch for all traits, and also sex effect for the BW1, BW8 and BW12 traits, and auxiliary variable: number of recording days for egg number); Z_1: incidence matrix that relate observations to the direct additive genetic effects of model; a: vector with direct genetic effects; e: vector of residual effects; Z_3: incidence matrix that relate observations to the maternal common environmental effects; c: vector of maternal common environmental effects; Z_2: incidence matrix that relate observations to the maternal

additive genetic effects of model; m: vector of maternal additive genetic effects; and Cov (a,m): covariance of direct and maternal additive genetic effects.

Results and discussion

In this study, the effectiveness of the selection process in the national breeding programme for ICs was evaluated by examining the extent of inbreeding and the resulting regression effects on some economic traits in ICBCs, using pedigree data from more than 400,000 ICs. The analysis of pedigree across the centres indicated that 61.6% of the ICs in Fars, 55.7% in Khorasan Razavi, 71.8% in Mazandaran, 78.8% in Esfahan, 78.5% in Yazd, and 88.7% in West Azerbaijan were found to be inbred. The average inbreeding coefficient varied from 2.2% in Fars to 6.3% in West Azerbaijan (Table 2).

Overall, in the populations examined, the number of inbred chickens was significant; however, the level of inbreeding was assessed to be below 10%, except in the most recent generations, where it exceeded 10%. The number of chickens with low inbreeding in these centres may be linked to inadequate data regarding parents or the execution of controlled matings within the population. In general, the frequency of inbred chickens with an inbreeding coefficient that exceeded 15% remained low across all centres. Close matings may be regarded as a potential factor contributing to the increased inbreeding coefficients.

The findings indicate that the inbreeding rate is increasing by less than 1% across the centres. Specifically, the average inbreeding rate is approximately 0.6% in Esfahan, West Azerbaijan, and Mazandaran; about 0.5% in Yazd and Khorasan Razavi; and around 0.3% in Fars. In an earlier study focused on Thai native chickens, the authors found that the inbreeding coefficient increased by 0.09% each

Table 2. Changes in the average inbreeding rates across generations. F, inbreeding coefficient; ge, average number of discrete generation equivalents.

Generation	Indigenous chicken breeding centres (ICBCs)											
	Mazandaran		Fars		Esfahan		West Azerbaijan		Yazd		Khorasan Razavi	
	F	(ge)	F	(ge)	F	(ge)	F	(ge)	F	(ge)	F	(ge)
1	0	0.96	0	1	0	0.5	0	0.5	-	-	0	1
2	0	1.47	0	1.49	0	0.61	0	0.74	0	1	0	2
3	0	1.74	0	1.75	0	0.82	0	1.70	0.005	1.99	0.001	3
4	0	1.81	0	1.14	0	0.53	0.004	2.63	0.009	2.98	0.005	4
5	0	1.90	0	2.11	0	1.05	0.004	3.45	0.010	3.99	0.006	5
6	0.001	2.91	0	2.05	0.001	2.06	0.004	4.23	0.005	4.97	0.016	6
7	0.004	3.58	0.001	3.05	0.005	3.06	0.031	5.15	0.024	5.98	0.018	6.9
8	0.003	3.73	0.002	4.03	0.002	4.07	0.042	6.20	0.015	6.96	0.020	8
9	0.004	4.56	0.007	5.02	0.012	5.07	0.032	7.17	0.023	7.96	0.026	9
10	0.008	5.76	0.012	5.99	0.019	6.07	0.046	8.17	0.030	8.97	0.037	10
11	0.006	6.66	0.027	6.98	0.029	7.08	0.050	9.05	0.040	9.97	0.050	10.99
12	0.029	7.72	0.035	7.88	0.030	8.07	0.055	9.95	0.049	10.97	0.058	12
13	0.030	8.71	0.011	5.89	0.039	9.01	0.068	10.72	0.053	11.97	-	-
14	0.050	9.74	0.024	7.12	0.045	10.04	0.060	11.59	0.058	12.97	-	-
15	0.051	10.75	0.020	8.08	0.055	11.07	0.065	12.46	0.067	13.97	-	-
16	0.053	11.75	0.022	9.05	0.064	12.07	0.068	13.16	-	-	-	-
17	0.054	12.76	0.036	10.05	0.075	13.07	0.080	14.14	-	-	-	-
18	0.079	13.75	0.031	11.06	0.081	14.07	0.089	15.19	-	-	-	-
19	0.068	14.76	0.034	12.07	0.093	15.07	0.096	16.16	-	-	-	-
20	0.077	15.68	0.041	13.06	0.098	16.07	0.119	17.16	-	-	-	-
21	0.089	16.76	0.049	14.06	0.114	17.07	0.128	17.97	-	-	-	-
22	0.097	17.76	0.051	15.07	-	-	-	-	-	-	-	-
23	0.108	18.76	0.058	16.06	-	-	-	-	-	-	-	-
24	0.116	19.76	0.061	17.06	-	-	-	-	-	-	-	-
25	0.123	20.76	0.067	18.07	-	-	-	-	-	-	-	-
26	0.141	21.76	-	-	-	-	-	-	-	-	-	-
Average	0.043	9.52	0.022	7.57	0.043	8.77	0.063	10.65	0.029	7.74	0.016	5.61
Average inbreeding coefficients in the inbreds	0.059		0.036		0.055		0.071		0.037		0.028	

year. They emphasized that this degree of inbreeding, at 0.09% per generation, falls within the commonly accepted threshold, which is typically regarded as approximately 1% per generation (Tongsiri *et al*, 2019). According to Nicholas (1989), the inbreeding rates of up to 0.5% per year should be acceptable in animal breeding programmes, since this results in a coefficient of variation in selection response of < 10% during the course of a 10-year selection period. A previous study estimated the inbreeding rate of Kokok Balenggek Chickens (KBC) under *ex situ* conservation as 0.31%. It was concluded that since the inbreeding rate was less than 1% – which implies that 1% of heterozygosity is lost per generation – the population is not at risk of extinction (Rusfidra *et al*, 2014). According to Weigel (2001), although the maintenance of genetic diversity and the preservation of selection response in future generations are vital for breeding programmes, the main effect of inbreeding at the farm level is inbreeding depression. Based on the estimated inbreeding rates in this study, the populations will retain their capacity to respond to selection in the future. Our results suggest that inbreeding rates are at a favourable level and selection has not adversely affected genetic diversity in Iranian ICBCs. This fact was also supported by genetic improvements made in the studied traits (Jelokhani-Niaraki and Ghorbani, in press). Nonetheless, this minimal level of inbreeding could potentially lead to a decline in the performance of traits.

Figure 1 depicts the trend of increased inbreeding across centres.

The inbreeding trend in all centres was almost identical. Although the studied populations are under genetic selection and closed, the populations in all centres exhibit a low inbreeding rate during first generations, which gradually increases with a gentle slope, ultimately reaching the highest inbreeding rates in the final generations. In the initial generations, the inbreeding rate was estimated to be zero, which may be attributed to the uncertainty surrounding pedigree information from those early generations. From another viewpoint, a pedigree with numerous unknown parents also contains several unknown common ancestors, and the obscurity of these common ancestors might be interpreted as their exclusion from the pedigree. Various studies have indicated that the accurate estimation of inbreeding is highly dependent on pedigree information. To accurately estimate the inbreeding coefficient in a population, two factors are particularly important: the completeness of the pedigree records and the careful control of mating. It has been observed that the amount of pedigree information used for estimating inbreeding has a direct effect on the inbreeding depression estimates. The increase in inbreeding is due to the fact that animals with similar breeding values are more likely to be related compared to those with different breeding values (Quinton *et al*, 1992; Miglior *et al*, 1995). According to a prior study regarding the effect of incomplete pedigrees on the estimates of inbreeding and inbreeding depression in Holstein and Jersey cows, the authors discovered that partial pedigrees result in a reduction of the average inbreeding estimate and the variance of these estimates within the groups of cows (Cassell *et al*, 2003). When attempting to perform a regression analysis of a response variable against inbreeding estimates derived from partial pedigrees, the resulting estimate of inbreeding depression may not correspond with that derived from complete pedigrees. It is logical to expect that a more comprehensive pedigree dataset will yield more precise assessments of inbreeding depression; however, various factors such as the precision of pedigree and phenotypic data, the size of datasets, statistical methods employed, and the models used will affect this result (Cassell *et al*, 2003). Besides the quality of the pedigree, the depth of the pedigree also plays a crucial role in more precisely estimating the inbreeding coefficient within the population. One method to assess and analyze the depth of the pedigree is to estimate the average number of discrete generation equivalents or the average number of generations known (ge). In our study, the assessment of this parameter revealed that the quality of the pedigree data in the ICBCs is satisfactory, and the inbreeding coefficients are estimated with a high degree of accuracy (Table 2).

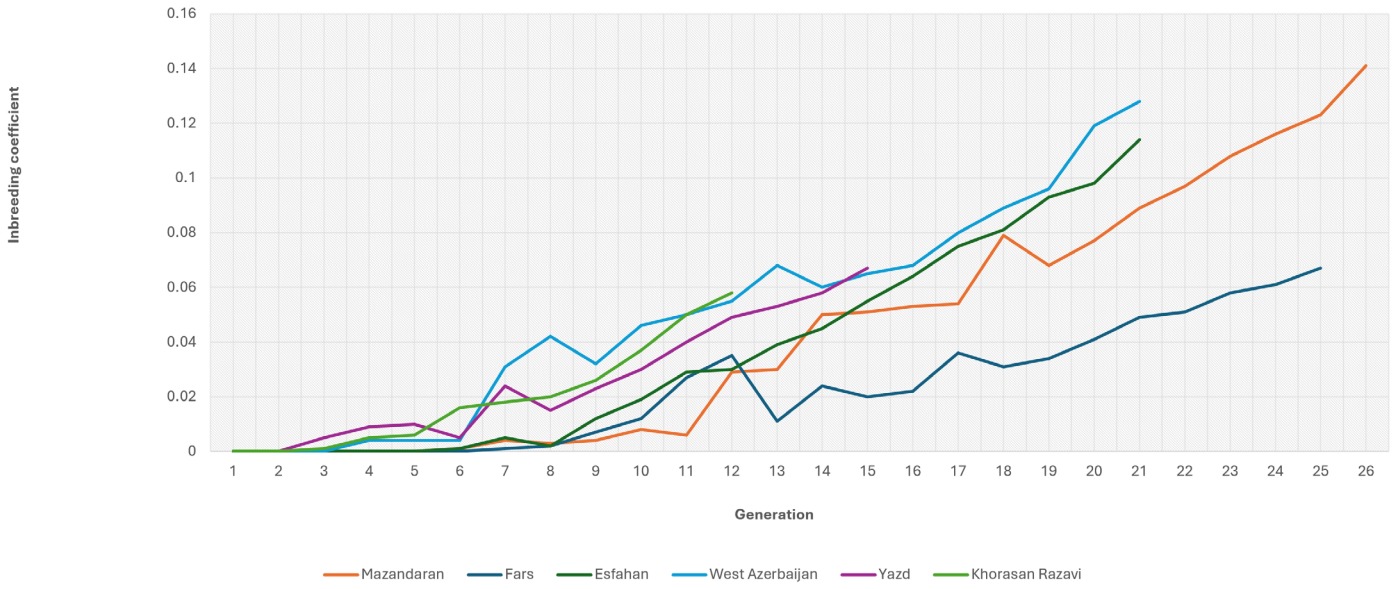


Figure 1. Trend of increasing inbreeding in indigenous chicken breeding centres (ICBCs)

To improve the accuracy of the most fitting model for predicting the breeding values of each trait, three different methodologies (LRT, AIC and BIC) were utilized. The findings from all centres indicated that there was no noticeable difference between the approaches used to select the most appropriate model for each trait. In a study that compared Legendre and B-spline random regression models for estimating the variance components of average birth weight in lambs per lambing in Mehraban sheep, the authors found that both AIC and BIC methods identified the model incorporating quadratic, linear, and quadratic B-spline fitted for random regressions and fixed regression, respectively, as the most appropriate model (Zamani *et al*, 2015). In another study regarding the modelling and fitting the growth curve model for Japanese quail under various nutritional conditions, the authors utilized both methods to fit three separate nonlinear growth models including Bertalanffy, Gompertz and Logistic. Their results revealed that both AIC and BIC methods identified the Logistic model as the most appropriate model (Dudusola *et al*, 2019).

The analysis of the regression coefficient of inbreeding on the traits revealed that the influence of inbreeding depression was low for most of the traits analyzed (Table 3).

Although inbreeding typically results in decreased fitness, the extent and specific effects of inbreeding can fluctuate considerably, influenced by the genetic makeup of the species or populations and their interactions with the environment (Hedrick and Kalinowski, 2000). In the present study, the significant effect of inbreeding on BW12 was found throughout all centres. This effect varied from -1.53 in Yazd to -3.51 in Esfahan. These values indicate that for each 1% rise in inbreeding, BW12 decreases by 1.53g in the Yazd population and by 3.51g in the Esfahan population. A negative trend was similarly identified for BW8, where every 1% increase in inbreeding leads to a reduction from 0.51g in the Khorasan Razavi population to 2.52g in the Esfahan population. The findings concerning ASM were largely consistent across all centres, indicating that for each 1% increase in inbreeding, ASM rises from 0.12 days in West Azerbaijan to 0.38 days in Fars. The findings from some studies regarding the effect of increased inbreeding on the ASM indicate that these effects differ across breeds. For example, in the Leghorn breed, increased inbreeding results in a rise in the ASM (Sewalem *et al*, 1999), whereas in the New Hampshire strain, it causes a decline in the ASM (Szwaczkowski *et al*, 2003). The results showed that the effects of inbreeding on egg-related traits, including the number of eggs, weight of the first egg, and average weight of eggs, were minimal. A study investigating the effects of inbreeding depression on body weight traits, average egg weight, age at first lay, and the percentage of fertile eggs in laying hen strains revealed minor effects (Szwaczkowski *et al*, 2004). In another study on native Thai chickens, the findings indicated that inbreeding had no effect on body weight traits, with the exception of the BW1 trait (Tongsiri *et al*, 2019). Ameli *et al* (1991) conducted a study on the cumulative inbreeding effects in commercial White Leghorn populations subjected to long-term reciprocal recurrent selection. After 23 generations, their findings revealed an annual inbreeding coefficient increase of 0.7% when full- and half-sibling matings were avoided. This progressive inbreeding was associated with a decline in egg production of three eggs annually. Additionally, another investigation examined the inbreeding depression associated with a 10% rise in inbreeding for different Leghorn lines and egg traits, including egg number, egg weight and egg mass weight. The values of inbreeding depression were found to be lower than the annual genetic progress values (Savas *et al*, 1999). A study performed on strains of white egg layers with the objective of evaluating the response to selection for fertility and hatchability, as well as the influence of inbreeding on these traits, demonstrated that inbreeding depression was not evident for any of the traits. This finding indicates that the selection offsets any adverse effects of inbreeding (Schmidt and Figueiredo, 2005). In cases where inbreeding does not rise too quickly, some evidences show that the resulting depression effect may be diminished through selection (Schmidt and Figueiredo, 2005; Gowe *et al*, 1993).

Conclusion

The findings from the pedigree analysis indicate that the inbreeding levels within the ICBCs are rising at a relatively gentle rate and remain within an acceptable range. Given the genetic improvements made in all six breeds, the inbreeding levels in these populations have been effectively managed, and it has not adversely affected the genetic progress of the population. As maintaining genetic diversity and minimizing inbreeding in the ICBCs is crucial for the success of breeding

Table 3. Regression coefficients of studied traits on inbreeding percentage. BW1, body weight at 1 day of age; BW8, body weight at 8 weeks of age; BW12, body weight at 12 weeks of age; ASM, age at sexual maturity; WSM, weight at sexual maturity; EN, egg number; EW1, egg weight at first day of laying and AEW, average egg weight at 28th, 30th and 32nd weeks. Significance level: $p < 0.01$. All regression coefficients reported in the table are statistically significant.

Traits	Indigenous chicken breeding centres (ICBCs)					
	Mazandaran	Fars	Esfahan	West Azerbaijan	Yazd	Khorasan Razavi
BW1	0.02	0.02	-0.008	-0.04	0.0003	-0.04
BW8	-1.32	-1.07	-2.52	-1.52	-0.69	-0.51
BW12	-2.04	-2.14	-3.51	-2.49	-1.53	-1.17
ASM	0.31	0.38	0.23	0.12	0.16	0.19
WSM	-0.11	0.27	-3.97	-1.22	-2.27	1.05
EN	-0.18	-0.07	-0.13	-0.01	-0.14	0.08
EW1	0.03	0.08	0.06	-0.01	-0.05	-0.05
AEW	-0.03	-0.01	0.02	-0.02	-0.05	-0.02

programmes, it can be inferred that the strategies for mating and the selection of superior chickens have been effectively implemented over the generations.

Acknowledgements

The authors would like to acknowledge the Deputy of Animal Production Affairs in the Ministry of Agriculture-Jahad for the financial support, as well as the indigenous chicken breeding centres for their contribution to data collection.

Author contributions

Saber Jelokhani-Niaraki: original draft preparation, writing, review and editing. Sholeh Ghorbani: methodology, investigation, data analysis, project administration, writing, review and editing.

Conflict of interest statement

The authors declare no conflict of interest.

Ethics statement

This study did not require review or approval by an ethics committee.

References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control* 19, 716-723. doi: https://doi.org/10.1109/TAC.1974.1100705

Ameli, H., Flock, D. K., Glodek, P. (1991). Cumulative inbreeding in commercial White Leghorn lines under long-term reciprocal recurrent selection. *Br. Poult. Sci* 32, 439-449. doi: https://doi.org/10.1080/00071669108417369

Barros, E. A., Brasil, L. D., Tejero, J. P., Delgado-Bermejo, J. V., Ribeiro, M. N. (2017). Population structure and genetic variability of the Segureña sheep breed through pedigree analysis and inbreeding effects on growth traits. *Small. Rumin. Res* 149, 128-133. doi: https://doi.org/10.1016/j.smallrumres.2017.02.009

Buranawit, K., Imboonta, N., Tongsiri, S., Masuda, Y., Phakdeedindan, P. (2025). Investigation of the effect of heat stress on egg production traits in Thai native chickens (Lueng Hang Kao Kabin) as determined by the temperature-humidity index. *Poult. Sci* 19, 105196. doi: https://doi.org/10.1016/j.psj.2025.105196

Cassell, B. G., Adamec, V., Pearson, R. E. (2003). Effect of incomplete pedigrees on estimates of inbreeding and inbreeding depression for days to first service and summit milk yield in Holsteins and Jerseys. *J. Dairy. Sci* 86, 2967-2976. doi: https://doi.org/10.3168/jds.S0022-0302(03)73894-6

D'ambrosio, J., Phocas, F., Haffray, P., Bestin, A., Brard-Fudulea, S., Poncet, C., Quillet, E., Dechamp, N., Fraslin, C., Charles, M., Dupont-Nivet, M. (2019). Genome-wide estimates of genetic diversity, inbreeding and effective size of experimental and commercial rainbow trout lines undergoing selective breeding. *Genet. Sel. Evol* 51, 1-15. doi: https://doi.org/10.1186/s12711-019-0468-4

Desta, T. T., Dessie, T., Bettridge, J., Lynch, S. E., Melese, K., Collins, M., Christley, R. M., Wigley, P., Kaiser, P., Terfa, Z., Mwacharo, J. M. (2013). Signature of artificial selection and ecological landscape on morphological structures of Ethiopian village chickens. *Animal. Genetic. Resources* 52, 17-29. doi: https://doi.org/10.1017/s2078633613000064

Dudusola, I. O., Oseni, S. O., Adeyemi, E. A. (2019). Modeling the growth curve of Japanese Quail under different nutritional environments. *Niger. J. Anim. Sci* 21, 53-58.

Fischer, T. M., Van der Werf, J. H. J., Banks, R. G., Ball, A. J. (2004). Description of lamb growth using random regression on field data. *Livest. Prod. Sci* 89, 175-185. doi: https://doi.org/10.1016/j.livprodsci.2004.02.004

Gowe, R. S., Fairfull, R. W., McMillan, I., Schmidt, G. S. (1993). A strategy for maintaining high fertility and hatchability in a multiple-trait egg stock selection program. *Poult. Sci* 72, 1433-1448. doi: https://doi.org/10.3382/ps.0721433

Harrison, S. (2017). Evaluating long-term direct and correlated selection response in White Plymouth Rock chickens selected for high or low 8-week body weight. MS.c thesis, University of Nebraska, *Lincoln*.

Hedrick, P. W., Kalinowski, S. T. (2000). Inbreeding depression in conservation biology. *Annu. Rev. Ecol. Syst* 31, 139-162. doi: https://doi.org/10.1146/annurev.ecolsys.31.1.139

IBM Corp. (2017). IBM SPSS Statistics for Windows, Version 25.0 [Computer software]. Armonk, NY: IBM Corp.

Jaturasitha, S., Kayan, A., Wicke, M. (2008). Carcass and meat characteristics of male chickens between Thai indigenous compared with improved layer breeds and their crossbred. *Arch. Anim. Breed* 51, 283-294. doi: https://doi.org/10.5194/AAB-51-283-2008

Jelokhani-Niaraki, S., Ghorbani, S., Esmailkhanian, S. (2021). Estimation of inbreeding depression for economic traits in Esfahan improved native chicken population. *J. Anim. Pro* 23, 313-324. doi: https://doi.org/10.22059/jap.2021.314496.623576

Jelokhani-Niaraki, S., Ghorbani, S. (in press). Evaluating the effectiveness of the long-term selective breeding program for indigenous chicken breeds. *Acta Sci. Anim. Sci.*

Lawal, R. A., Al-Atiyat, R. M., Aljumaah, R. S., Silva, P., Mwacharo, J. M., Hanotte, O. (2018). Whole-genome resequencing of red jungle fowl and indigenous village chicken reveal new insights on the genome dynamics of the species. *Front. Genet* 9: 264. doi: https://doi.org/10.3389/fgene.2018.00264

Lawal, R. A., Hanotte, O. (2021). Domestic chicken diversity: Origin, distribution, and adaptation. *Anim. Genet* 52, 385-394. doi: https://doi.org/10.1111/age.13091

Lewis, F., Butler, A., Gilbert, L. (2011). A unified approach to model selection using the likelihood ratio test. *Methods. Ecol. Evol* 2, 155-162. doi: https://doi.org/10.1111/j.2041-210X.2010.00063.x

Magothe, T. M., Okeno, T. O., Muhuyi, W. B., Kahi, A. K. (2012). Indigenous chicken production in Kenya: II. Prospects for research and development. World's. *Poult. Sci. J* 68, 133-144. doi: https://doi.org/10.1017/S004393391200013X

Meyer, K. (2007). WOMBAT, A tool for mixed model analyses in quantitative genetics by REML. *J. Zhejiang. Univ. Sci*. B 8, 815-821. doi: https://doi.org/10.1631/jzus.2007.B0815

Miglior, F., Burnside, E. B., Dekkers, J. C. (1995). Nonadditive genetic effects and inbreeding depression for somatic cell counts of Holstein cattle. *J. Dairy. Sci* 78, 1168-1173. doi: https://doi.org/10.3168/jds.S0022-0302(95)76734-0

Milkias, M., Molla, M., Tilahun, S. (2019). Productive and reproductive performance of indigenous chickens in Gena Bossa District of Dawro Zone, Ethiopia. *Int. J. Livest. Prod* 10, 24-32. doi: https://doi.org/10.5897/IJLP2018.0551

Nicholas, F.W. (1989). Incorporation of new reproductive technology in genetic improvement program. In *Evolution*

and Animal Breeding. ed. Hill WG and Mackay TFC (Wallingford: CAB International). 203-209.

Padhi, M. K. (2016). Importance of indigenous breeds of chicken for rural economy and their improvements for higher production performance. *Scientifica (Cairo)* 2604685. doi: https://doi.org/10.1155/2016/2604685

Quinton, M., Smith, C., Goddard, M. E. (1992). Comparison of selection methods at the same level of inbreeding. *J. Anim. Sci* 70: 1060-1067. doi: https://doi.org/10.2527/1992.7041060x

Radwan, L. M. (2020). Genetic improvement of egg laying traits in Fayoumi chickens bred under conditions of heat stress through selection and gene expression studies. *J. Therm. Biol* 89, 102546. doi: https://doi.org/10.1016/j.jtherbio.2020.102546

Rusfidra., Tan Marajo, S. D., Heryandi, Y., Oktaveriza, B. (2014). Estimation of inbreeding rate in Kokok Balenggek Chicken (KBC) population under ex-situ conservation. *Int. J. Poult. Sci* 13, 364-367. doi: https://doi.org/10.3923/ijps.2014.364.367

Sargolzaei, M., Iwaisaki, H., Colleau, J. J. (2006). A tool for monitoring genetic diversity. In proceedings of the 8th World Congress Genetics Applied Livestock, ProBelo Horizonte

Savas, T., Preisinger, R., Röhe, R., Kalm, E., Flock, D. K. (1999). Auswirkungen der Inzucht auf Leistungsmerkmale und deren genetische Parameter bei Legehennen. *Arch. Fur. Geflugelkd* 63, 246-251.

Schmidt, G. S., Figueiredo, E. A. P. (2005). Selection for reproductive traits in white egg stock using independent culling levels. *Braz. J. Poult. Sci* 7, 231-235.

Schwarz, G. (1978). Estimating the dimension of a model. *Ann. Stat* 6, 461-464. doi: https://doi.org/10.1214/aos/1176344136

Selvaggi, M., Dario, C., Peretti, V., Ciotola, F., Carnicella, D., Dario, M. (2010). Inbreeding depression in Leccese sheep. *Small. Rumin. Res* 89, 42-46. doi: https://doi.org/10.1016/j.smallrumres.2009.12.005

Sewalem, A., Johansson, K., Wilhelmson, M., Lillpers, K. (1999). Inbreeding and inbreeding depression on reproduction and production traits of White Leghorn lines selected for egg production traits. *Br. Poult. Sci* 40, 203-208. doi: https://doi.org/10.1080/00071669987601

Siebenmorgen, C., Mörlein, J., Strack, M., Tetens, J., Mörlein, D. (2024). Enhancing agro-biodiversity in chicken: a sensory comparison of broths from German local chicken breeds and their crossbreeds. *Poult. Sci* 103, 103683. doi: https://doi.org/10.1016/j.psj.2024.103683

Stefanetti, V., Mancinelli, A. C., Pascucci, L., Menchetti, L., Castellini, C., Mugnai, C., Fiorilla, E., Miniscalco, B., Chiattelli, D., Franciosini, M. P., Proietti, P. C. (2023). Effect of rearing systems on immune status, stress parameters, intestinal morphology, and mortality in conventional and local chicken breeds. *Poult. Sci* 102, 103110. doi: https://doi.org/10.1016/j.psj.2023.103110

Szalay, I. T., Lan Phuong, T. N., Barta, I., Bodi, L., Emodi, A., Szentes, K. A., Dong Xuan, K. D. T. (2016). Conservation aspects of meat producing ability and heterosis in crosses of two natively different local Hungarian chicken breeds. *Int. J. Poult. Sci* 15, 442-447. doi: https://doi.org/10.3923/ijps.2016.442.447

Szwaczkowski, T., Cywa-Benko, K., Wezyk, S. (2003). A note on inbreeding effect on productive and reproductive traits in laying hens. *Anim. Sci. Pap.* Rep 21, 121-129.

Szwaczkowski, T., Cywa-Benko, K., Wezyk, S. (2004). Curvilinear inbreeding effects on some performance traits in laying hens. *J. Appl. Genet* 45, 343-345.

Tolone, M., Sardina, M. T., Criscione, A., Lasagna, E., Senczuk, G., Rizzuto, I., Riggio, S., Moscarelli, A., Macaluso, V., Di Gerlando, R., Cassandro, M., Portolano, B., Mastrangelo, S. (2023). High-density single nucleotide polymorphism markers reveal the population structure of 2 local chicken genetic resources. *Poult. Sci* 102, 102692. doi: https://doi.org/10.1016/j.psj.2023.102692

Tongsiri, S., Jeyaruban, G. M., Hermesch, S., Van Der Werf, J. H. J., Li, L., Chormai, T. (2019). Genetic parameters and inbreeding effects for production traits of Thai native chickens. *Asian-Australas. J. Anim. Sci* 32, 930-938. doi: https://doi.org/10.5713/ajas.18.0690

Van Wyk, J. B., Fair, M. D., Cloete, S. W. P. (2009). Case study: the effect of inbreeding on the production and reproduction traits in the Elsenburg Dormer sheep stud. *Livest. Sci* 120, 218-224. doi: https://doi.org/10.1016/j.livsci.2006.10.005

Weigel, K. A. (2001). Controlling inbreeding in modern breeding programs. *J. Dairy. Sci* 84, E177-184. doi: https://doi.org/10.3168/jds.S0022-0302(01)70213-5

Yadav, A., Jain, A., Sahu, J., Dubey, A., Gadpayle, R., Barwa, D. K., Kumar, V. (2019). A review on the concept of inbreeding and its impact on livestock. *Int. J. Fauna. Biol. Stud* 6, 23-30.

Zamani, P., Amirabadi-Farahani, M., Aliarabi, H., Malecky, M. (2016). Comparison of different Legendre and B-Spline random regression models to estimate variance components for average birth weight per lambing in Mehraban sheep. *Iran. J. Anim. Sci* 46, 407-415. doi: https://doi.org/10.22059/ijas.2015.56826

Genetic Resources

ORIGINAL ARTICLE

Genetic Resources (2026), 7 (13), 65–76
DOI: 10.46265/genresj.DOAE5566
https://www.genresj.org
ISSN: 2708-3764

The local crop varieties (farmers' varieties) registration system in Nepal: Past, present and future

Bal Krishna Joshi[a,*], Pradip Thapa[a], Benu Prasai[b] and Dila Ram Bhandari[b]

[a] *National Agriculture Genetic Resources Center (Genebank), NARC, Khumaltar, Nepal*
[b] *Seed Quality Control Center (SQCC), MoALD, Kathmandu, Nepal*

Abstract: Farmers' varieties, characterized by their broad genetic base and ecological resilience, were historically not recognized by Nepal's formal seed system. However, after decades of research, policy dialogue and advocacy efforts, the Seed Regulation of 2013 introduced a provision allowing farming communities to register their varieties through a simplified process. This milestone was further reinforced by the Seed Act of 2022 and the Seed Regulation of 2024, which legally defined native and local landraces, ensuring their formal recognition and protection. Despite these advancements, market opportunities for site-specific landraces remain limited, posing a challenge for widespread commercialization. Nevertheless, the registration system has significantly contributed to Nepal's seed sovereignty, empowering farmers economically, promoting agrobiodiversity conservation and enhancing agricultural sustainability. By integrating formal recognition, economic incentives and ecological resilience, Nepal is fostering a farmer-driven, sustainable agricultural future where local varieties play a crucial role in food, nutrition, health and business security, climate adaptation and rural livelihoods.

Keywords: Formal seed system, landrace, seed cycle, farmers' rights, farmer-managed seed system

Citation: Joshi, B. K., Thapa, P., Prasai, B. and Bhandari, D. R. (2026) "The local crop varieties (farmers' varieties) registration system in Nepal: Past, present and future", *Genetic Resources*, 7(13), pp. 65–76. doi: 10.46265/genresj.DOAE5566.

Introduction

Nepal is home to approximately 30,000 site-specific crop landraces cultivated across diverse agro-ecological zones, ranging from 60 to 4,700 meters in altitude (Joshi et al, 2020b). Despite this rich genetic diversity, the country remains heavily dependent on foreign germplasms, with 95–100% of its breeding materials sourced from outside (Chaudhary et al, 2016). Only 5% of Nepal's native agricultural genetic resources (AGRs) have been utilized in research, and just 37 local landraces of 19 crops have contributed to the development of 41 modern varieties (SQCC, 2024; Joshi 2017). This underutilization of indigenous genetic resources highlights a significant gap in Nepal's agricultural policy and breeding programmes.

Among the six components of AGRs – crops, livestock, forage, agro-insects, agro-microbes, and aquatic – the formal seed system in Nepal only accounts for crops and forages, while farmers continue to maintain and produce all six components (Joshi et al, 2020b). The country's seed system can be categorized into three broad types based on breeding and legal status: the formal seed system, the non-formal seed system (privately managed system of improved/ exotic varieties without registration), and the informal seed system (traditional farmer-managed system of landraces without formal regulation) (Gurung et al, 2020; Joshi et al, 2020d;

* Corresponding author: Bal Krishna Joshi (joshibalak@yahoo.com)

Received: 01.03.2025 | Accepted: 14.01.2026 | Published online: 17.02.2026

LI-BIRD and the Development Fund, 2017). However, with the expansion of the formal seed system, farmers have been largely excluded from the seed registration process, limiting their rights over the seed production cycle (De Jonge et al, 2025). Despite this, the majority of seed transactions in Nepal still occur through informal channels (Gurung et al, 2020; Joshi et al, 2020d).

In Nepal's formal system, crop varieties can be listed/notified through two mechanisms: registration (in which one-season evaluation data is enough) and release (in which three-season evaluation data is required) (MoAD, 2013; Joshi et al, 2017b). As of September 2025, 956 varieties of 85 different crops had been developed and notified in the national seed board (SQCC, 2024). Various types of cultivars exist within the country, including registered and released (R&R) varieties, native landraces, local varieties, hybrids, open-pollinated varieties (OPVs), exotic varieties, cultivar mixtures, evolutionary populations, composites, and pure lines. Farmers' varieties are broadly classified into two types: native or indigenous landraces and local varieties or local landraces (MoALD, 2024); both are called farmers' varieties. A native or indigenous landrace refers to a genotype that has evolved or originated in a specific area, with at least one unique trait or gene that developed in that region and has been traditionally cultivated there. A landrace is a traditional crop population naturally adapted to a local area, not bred by formal breeders, while a variety is a scientifically bred and improved crop developed by a professional plant breeder. A local variety and landrace is a crop that has been fully adapted to a particular area, experiencing zero environmental shock, or any variety that existed in Nepal before 1951 and has been continuously cultivated in a specific region for over 60 years.

Although genetic diversity is key to agricultural sustainability and climate resilience, the formal registration system prioritizes uniformity through distinctness, uniformity and stability (DUS) testing – a criterion designed for commercial breeding rather than for recognizing the diverse nature of farmer-managed landraces (MoALD, 2024). Nepal's seed regulatory framework requires DUS testing for formal variety registration, which benefits seed companies and breeders but not traditional farmers (MoA, 1997). This restriction further discourages on-farm conservation and the formal recognition of native landraces, undermining their role in sustainable agriculture. Given the increasing challenges posed by climate change and shifting agricultural landscapes, Nepal must balance a localized seed system with a globalized product system to ensure both food security and economic sustainability. Farmers are not just cultivators but also breeders, innovators, and custodians of genetic diversity and traditional knowledge (Gauchan et al, 2020b). They play a critical role as primary sources of agricultural practices, germplasm conservation and sustainable farming techniques.

While formal plant breeders require DUS criteria to identify a variety, farmers recognize and maintain their landraces even with inherent diversity. These landraces, though genetically diverse, retain their identity and functional traits over generations, even in cross-pollinated and open-field conditions. Farmers have developed sophisticated selection and management practices, ensuring that their traditional varieties remain stable and locally adapted. The ability of farmers to maintain the genetic integrity of landraces across different agroecological sites is evidence of the effectiveness of farmer-led conservation strategies.

One of the most effective conservation-through-use approaches is the formal registration of native landraces, which provides legal recognition, enabling farmers to produce, market and benefit from their seeds while ensuring their continued conservation. This approach bridges the gap between informal and formal seed systems, promoting seed sovereignty, economic incentives for farmers, and long-term sustainability of genetic resources. By acknowledging and registering farmers' varieties, Nepal can strengthen on-farm conservation efforts, enhance climate resilience and safeguard agricultural biodiversity for future generations.

To support farmer-led conservation and innovation, Nepal's agricultural policies must transition from merely formalizing agricultural inputs – such as seeds, pesticides and fertilizers – to formalizing farmer-produced outputs and ensuring market guarantee of each product. Farmers should have full rights over their seeds, traditional knowledge and genetic resources, with an equitable access and benefit-sharing mechanism in place. Recognizing the urgency of agrobiodiversity conservation, this paper examines the evolution of Nepal's local variety registration system, explores its present challenges, and envisions a future framework that prioritizes farmer empowerment, sustainability and the legal recognition of diverse genetic resources.

Methodology

This study employed a comprehensive approach integrating review, observation, survey and experience (ROSE) to analyze the farmers' variety registration system in Nepal from 2000 to 2024. The methodology is structured as follows:

1. Review of literature and documents: A thorough review of relevant literature, policy documents, reports and Community Biodiversity Registers (CBRs) was conducted to understand the historical and policy evolution of local variety registration. Key sources included national and international publications, government policies, legal frameworks and institutional reports related to farmers' varieties and agrobiodiversity conservation
2. Field observations: Multiple observations were carried out over the years in various settings, including farmers' fields, agricultural fairs, seed storage facilities, community seed banks and the Nepal (National) Genebank. These observations provided firsthand insights into traditional seed-saving practices, varietal diversity, farmers' perceptions, and the challenges in the registration process
3. Surveys and stakeholder consultations: To gather field-level data and stakeholder perspectives, the study employed 35 focus group discussions (FGDs) with farmers, seed custodians and community-based organizations; 100 key informant interviews (KIIs) involving policymakers, researchers, extension workers and representatives from seed regulatory bodies; 50+ telephone surveys to supplement and verify information from stakeholders across different regions.
4. Documentation of authors' experiences: The experiences of all authors over the study period (2000–2024) were systematically documented, incorporating insights from field engagements, institutional involvement and direct participation in seed registration processes. Challenges faced, major policy shifts, significant events and scientific

evidence were synthesized to provide a well-rounded perspective.

This multi-method approach ensured a holistic and evidence-based analysis of the local variety registration system in Nepal, capturing both historical trends and future trajectories. To generate empirical evidence, numerous on-farm trials were conducted to evaluate the performance, adaptability and resilience of different local varieties in varying agroecological conditions. Additionally, the pedigree analysis of many released varieties was carried out to trace their genetic lineage and understand the contributions of farmers' varieties in modern breeding programmes. Both farmers' varieties and improved varieties were assessed for their genetic diversity, adaptability and farmer preference, providing crucial insights into conservation needs and utilization strategies. A policy gap analysis (mainly through policy document review, implementation status and strategy study, and interaction with farmers, experts and seed companies) was conducted to evaluate the effectiveness of existing legal frameworks and institutional mechanisms for local variety registration. Survey reports were developed based on focus group discussions, key informant interviews and telephone surveys to capture stakeholders' perspectives. The drivers of genetic erosion of native landraces were studied in depth, with efforts made to estimate the percentage of loss and underlying factors contributing to the decline in agrobiodiversity.

Several CBRs maintained by the community were analyzed to document the status and dynamics of local varieties at the community level. A farmers' varietal catalogue (Gurung et al, 2019) of ten communities was developed to systematically document traditional seed diversity, their agronomic traits and their historical significance. To facilitate participatory research and conservation, Five-Cell Analysis, a method (Joshi et al, 2022) used to assess varietal richness and distribution, was conducted in more than ten sites during 2000 to 2012.

The study also incorporated sensory evaluations to assess farmers' and consumers' preferences for specific landraces based on taste, texture, and other quality attributes. Community Seed Banks (CSBs) were established and documented, highlighting their role in the conservation and distribution of local varieties. In addition, nutritional analysis was performed to determine the health benefits of various landraces, and product diversification efforts were undertaken to explore new ways of utilizing traditional crops for economic and food security benefits.

Results and discussion

Historical lessons and conventional practices

Nepal's past agricultural policies and practices have largely emphasized the promotion of monogenotypes, favoring single, homogeneous and uniform exotic varieties over diverse native landraces (Gauchan et al, 2017; Joshi 2017). This approach led to large-scale cultivation of a few improved varieties, particularly in staple crops like rice, wheat and maize (SQCC, 2024). While these crops have been prioritized in research, education and development, 85% of Nepal's native agrobiodiversity consists of neglected and underutilized species (NUS) and future/smart crops (F/SC), (Joshi et al, 2020b) which have received little attention in research, development and policy frameworks. As a result, many traditional crops, which are highly nutritious, healthy and climate-resilient, remain underutilized (Joshi, 2017).

The heavy dependence on exotic germplasm and technologies has contributed to the erosion of native crop biodiversity (Chaudhary et al, 2016). Studies indicate that 50–100% of crop biodiversity has been lost in different regions of Nepal, depending on crop type and location (Joshi et al, 2020b).

The existing agricultural policies have also reinforced inequality in access to incentives and services. Only R&R varieties are eligible for government incentives, subsidies and extension support, leaving farmers' varieties at a disadvantage (Gauchan et al, 2017).

Additionally, past policies have promoted a high seed replacement rate, which prioritizes frequent replacement of traditional varieties with modern hybrids and varieties. While this may enhance yield in the short term, it disrupts traditional seed cycles, making farmers completely dependent on external seed sources and accelerating the loss of indigenous landraces (Gurung et al, 2020; Joshi et al, 2020c). This dependency has gradually eroded farmers' rights (e.g. germplasm as private property, registration provision by an individual farmer, seed saving, marketing, etc.), turning them into mere consumers of commercial seed companies, despite the fact that farmers produce many essential agricultural inputs beyond edible crops.

Farmers' varieties, also known as native landraces, are an invaluable component of Nepal's agricultural heritage. These varieties possess rich genetic diversity, which plays a crucial role in maintaining climate-resilient and sustainable agricultural systems (Joshi et al, 2018; 2023a). The greater the genetic diversity, the stronger the adaptive capacity of crops to environmental changes, pests and diseases, ensuring the long-term sustainability of farming (Zhu et al, 2000; Neupane et al, 2023). However, despite their importance, the genetic diversity of landraces has not been fully recognized in national policies, limiting their integration into formal seed systems (Joshi, 2017; MoA, 1997). This lack of policy recognition contradicts the fundamental need for genetic diversity in resilient agriculture.

Scientific and empirical evidences for policy reformulation

Between 2000 and 2024, extensive research, field trials, policy analyses and participatory approaches have generated both empirical and scientific evidence supporting the advantages of farmers' varieties over improved varieties (Joshi et al, 2023a). Numerous on-farm trials demonstrated that farmers' varieties perform better in terms of ecological yield, taste, adaptability, genetic diversity, farmer preferences, risk-bearing capacity and resilience across different agroecological conditions (Shrestha et al, 2013; Sthapit, 2013). Unlike modern hybrids, which often require high external inputs, farmers' varieties have shown higher adaptability to local environments, making them more sustainable for smallholder farmers (Bajracharya et al, 2012; Poudel et al, 2015; Bhandari et al, 2017; Gauchan et al, 2018; 2020a; Shrestha and Rana, 2018; Gurung et al, 2020; Joshi et al, 2020a; Karkee et al, 2023).

Pedigree analysis of released varieties in Nepal revealed

that most of the parent materials used for breeding improved varieties originated from outside the country, emphasizing the reliance on exotic germplasm (Chaudhary et al, 2016; Joshi, 2017). Comparative studies between farmers' varieties and uniform, exotic varieties highlighted key advantages of local landraces, including higher genetic variation, better taste, climate resilience, low input requirements, nutritional richness and multiple ecological benefits (Joshi et al, 2018; 2023a; Sthapit et al, 2019). Moreover, farmers' varieties exhibited a high evolutionary rate, making them more adaptive and sustainable in the face of climate change. These findings underscored major policy gaps, as policies before 2013 primarily favoured improved and uniform varieties, often excluding farmers' varieties from formal recognition and incentives (MoA, 1997).

One major reason for the decline of farmers' varieties was the aggressive promotion of R&R varieties, which were often incentivized by government policies (Upreti and Upreti, 2002; Joshi et al, 2020b). Farmers have reported a 50–100% loss of traditional landraces, largely attributed to the widespread adoption of modern crop varieties (Joshi et al, 2020d). Farmers reported total failures in some improved varieties due to their lack of adaptability, while complete crop failure was rarely observed in farmers' varieties. The high genetic diversity within farmers' varieties allowed them to remain more dynamic, evolving and resilient to climate change and pest outbreaks (Joshi et al, 2023a).

Farmers' varieties also play a crucial role in ensuring food security, nutrition security, health security, business security and environmental sustainability. For example, landraces of rice, maize and wheat demonstrated yields 20% higher than the national average, while minor crops outperformed improved varieties by 60% (NAGRC, 2024). Additionally, 100% of indigenous varieties still contribute to 83% of total cultivated crops, showing their continued significance in Nepalese agriculture (Gauchan et al, 2020b; Gurung et al, 2020). While modern varieties had higher carbohydrate content, landraces were richer in essential micronutrients, phytochemicals and antioxidants, promoting better nutrition security (Joshi et al, 2020d). Furthermore, since indigenous varieties are grown agroecologically, they contribute to health security by reducing exposure to chemical residues. Their higher market value (Gauchan et al, 2020a), better taste, and balanced nutrient composition provide business security to farmers. Finally, as climate-resilient, organically grown crops with high ecological yields, farmers' varieties contribute to environmental sustainability and low-risk farming systems in Nepal.

Milestones in farmers' variety registration

The formal discussion on the comparative advantages of farmers' varieties over improved R&R varieties began in 2000, marking the initial step towards their recognition in Nepal's seed system (Table 1). Around this time, Nepal Genebank, Local Initiatives for Biodiversity, Research, and Development (LI-BIRD), and Bioversity International started working collaboratively on policy dialogues, policy gap analysis, and evidence generation to advocate for the importance of farmers' varieties. Through continuous efforts, the unique advantages and diversity of farmer-managed varieties were explored (through on-farm trials, focus group discussions and interactions with farmers), leading to increased awareness among policymakers, researchers and farming communities.

To provide scientific validation and support for farmers' varieties registration, participatory on-farm trials were initiated across different districts (Sthapit and Jarvis, 2003). Community-based initiatives played a significant role in generating both scientific and local knowledge. Diversity fairs were organized to showcase and promote farmer-managed seed diversity, while Diversity Field Schools facilitated participatory learning and seed selection processes among farming communities. Participatory landrace enhancement programmes were carried out to improve the productivity and resilience of selected landraces through farmer-led breeding efforts. Diversity blocks were established in different regions to conserve and monitor traditional varieties *in situ*, ensuring their continued use and adaptation.

To facilitate the recognition and registration of local varieties, various strategies and policy documents were developed to convince the relevant authorities. Several policy dialogues, travelling seminars and field visits were organized to bring together policymakers, researchers, farmers and seed experts at local, provincial and national levels. Regular interaction programmes were conducted with farmers and breeders to understand their concerns and integrate their perspectives into policy discussions. Additionally, workshops, sharingshops and writeshops were held to draft policy recommendations, ensuring a participatory approach to decision-making. Continuous engagement with authorities through publication sharing and organoleptic tests further supported evidence-based policy advocacy.

The initial format, which was simple and accessible online for farmers' variety registration, was developed by key breeders from Nepal Genebank and the Seed Quality Control Centre (SQCC), ensuring it was farmer-friendly and easy for farming communities to fill out independently (MoAD, 2013). Recognizing the importance of traditional knowledge in the registration process, the format was later revised to include detailed information on origin, cropping history, agromorphological traits, social significance and economic values of farmers' varieties (MoALD, 2024). This formal format was prepared through a collaborative effort involving plant breeders, seed experts, conservationists and farmers, ensuring that the format was practical, inclusive and scientifically sound. To strengthen institutional and community-level capacities, orientation programmes, training sessions and varietal proposal writeshops were conducted for farmers, researchers and policymakers. Hands-on training in on-farm trials, data recording, and participatory variety selection enabled farmers and researchers to generate robust scientific data to support variety registration.

A significant milestone was the development of detailed guidelines for the registration, production, maintenance and distribution of farmers' varieties in 2022 by Nepal Genebank, LI-BIRD and CSBs. Special attention was given to maintaining genetic diversity during seed multiplication, reinforcing the importance of conserving the rich agrobiodiversity of farmers' varieties while promoting their sustainable use and commercialization. This not only enhanced farmers' technical capacities but also provided ownership rights to farmer groups, empowering them to maintain, produce and sell their own seeds legally. Unlike in the past, where the formal seed system detached farmers from the seed cycle, this initiative allowed farmers to regain control over seed management (Gauchan et al, 2017).

To ensure the conservation of original genetic materials, passport data and the original seed lots of registered varieties are maintained by the Nepal Genebank under the Nepal Agricultural Research Council (NARC) (NAGRC, 2024). However, despite these achievements, the registration process in Nepal has so far been limited to orthodox seeds (seeds that can be dried and stored for long periods) (SQCC, 2024). The system still lacks a mechanism for registering non-orthodox seeds, including vegetatively propagated crops, recalcitrant seeds and cultivar mixtures, highlighting the need for further advancements in the farmers' variety registration framework.

The formal registration of native landraces in Nepal began in 1979, when plant breeders started registering traditional varieties after selection (Figure 1) (Sthapit et al, 1998; SQCC, 2024). However, informal registration of farmers' varieties had started in 1986 under the Plant Genetic Resource Unit (Seed Bank) of the Agriculture Botany Division. Despite these early efforts, until 2013, only formal plant breeders and research institutes had the authority to register varieties, limiting farmers' ability to legally recognize and commercialize their own traditional seeds (MoA, 1997). A major shift occurred in 2013, when Nepal's Seed Regulation was amended, allowing groups of farmers to register their landraces in the National Seed Board (NSB) as their own business items (MoAD, 2013). This revision included Schedule D, a simplified format for the registration and release of farmers' varieties, officially recognizing farmers as breeders. This paved the way for the first-ever formal registration of a farmers' variety in 2014. The first registered farmers' variety was broad leaf mustard (Gujmuje Rayo) from Dalchoki, Lalitpur, facilitated by the Dalchoki Community Seed Bank, Nepal Genebank, and SAHAS-Nepal (SQCC, 2024).

The registration process for Gujmuje Rayo and another landrace, Dunde Rayo, involved on-farm trials conducted in Dalchoki, Lalitpur, where all released varieties of broad leaf mustard were tested alongside the farmers' varieties. The trial was evaluated by experts, farmers, members of the Variety Approval Release and Registration Sub-Committee (VARRSC), the Nepal Genebank team, and SAHAS-Nepal representatives. A travelling seminar and organoleptic test were conducted, where all tested varieties were cooked and evaluated for taste, texture and other consumer preferences. Additionally, simple agromorphological data were collected to support the registration proposal. With technical and institutional support from Nepal Genebank and SAHAS-Nepal, the Dalchoki Community Seed Bank formally submitted a proposal to the NSB for the registration of these varieties. After evaluation, the VARRSC approved the registration under Schedule-D of the Seed Regulation 2013, officially recognizing Gujmuje Rayo as the first registered farmers' variety in Nepal. This milestone set an important precedent for farmer-led conservation, recognition and commercialization of native

Table 1. Major events in the process of registration of farmers' varieties in Nepal. Source: Sthapit et al, 1998; Joshi and Witcombe, 2003; Subedi et al, 2005; MoAD, 2013; Chaudhary et al, 2016; Bhandari et al, 2017; Gauchan et al, 2017; Joshi, 2017; Joshi et al, 2017a; Thapa et al, 2019; Joshi et al, 2020b; MoALD, 2022; MoALD, 2024; SQCC, 2024.

Year	Major actions	Organizations involved	Remarks
2000–2012	Many round discussions about legal provision for native landraces (e.g. meeting, workshop, travelling seminar, interaction meeting), evidence generation, history and gap analysis, on-farm trial for native landraces	Nepal Genebank, LI-BIRD, Bioversity International	Carried out these actions in many districts
2012	Accessioning system of Genebank for submitted varieties in Variety Approval, Release and Registration Sub-Committee (VARRSC) coordinated by the Seed Quality Control Centre (SQCC)	Nepal Genebank	Nepal Genebank started providing accession numbers and seeds of such varieties are maintained in Nepal Genebank. Only for orthodox seeds
2013	Legal provision of registration of native and local landraces of crops (heterogenous varieties): Seed Regulation 2013, Rule-12, By-rule- 2: Annex-1 Schedule (Clause)-D	SQCC, Nepal Genebank	Very simple format, at first prepared and made available online
2018	Format revision, verification and testing during 2nd National Workshop of Community Seed Banks	SQCC, Nepal Genebank and LI-BIRD, Community Seed Banks	Simple existing format elaborated and shared among farmers and breeders
2019	Consultation workshop on farmers' variety registration and commercialization; training workshop on local variety registration process	Nepal Genebank, LI-BIRD, Community Seed Banks	Mutual understanding among farmers, seed experts and breeders
2021	A training workshop on farmers' variety registration proposal development	Nepal Genebank, LI-BIRD, Community Seed Banks	Practical exercise on format of registration
2022	Training, workshop and writeshop for drafting detail guideline on farmers' variety registration and maintenance, A national workshop on native varieties registration in Nepal: Issues, Achievements and Challenges	Nepal Genebank, LI-BIRD, Community Seed Banks, SQCC	Issues of registration and maintenance identified and discussed
2022	Seed Act 1988 (2nd amendment)	SQCC	Farmers' variety recognized
2023	Details guidelines preparation	Nepal Genebank, LI-BIRD	Seed production, maintenance, labelling and marketing
2024	Seed regulation (revised)	SQCC	Details guidelines and defines native and local landraces

landraces, strengthening the integration of traditional seed systems into formal seed governance.

Creating an enabling environment

In this process, multiple stakeholders – including Community Seed Banks and farmers – have been actively engaged. Creating an enabling environment is vital to support the revision of policies. Key mechanisms include regular interaction, field visits, meetings, workshops, on-farm trials, travelling seminars, and training. Among these, two components stand out as especially important: capacity enhancement and policy engagement and advocacy.

Regular capacity-building programmes (training workshops, exchange visits, practical exercises, travelling seminars, writeshops and interactive meetings at local, regional and national levels) were conducted for farmers and relevant stakeholders to strengthen their understanding and skills in various aspects of farmers' variety registration and conservation in more than 25 districts. Three organizations, namely Nepal Genebank, LI-BIRD and Bioversity International, were directly involved in capacity building. These programmes focused on the values and importance of farmers' varieties, emphasizing their role in genetic diversity, climate resilience and sustainable agriculture (Gauchan et al, 2016; Gauchan et al, 2020a; Joshi et al, 2020a; 2020b; 2020c). Training sessions covered participatory plant breeding, data generation and documentation, ensuring that farmers could scientifically record and validate their varieties. Additionally, farmers and local stakeholders were trained in PowerPoint preparation and presentation skills, enabling them to effectively communicate their experiences and research findings. These activities fostered collaboration among farmers, researchers, policymakers and conservationists, ultimately strengthening the farmer-led seed system and promoting farmers' rights and agrobiodiversity conservation in Nepal.

The evidence generated on the benefits of farmers' varieties was shared, presented and discussed with various stakeholders and key authorities (e.g. MoALD, SQCC, NARC, DoA, LI-BIRD, Institute of Agriculture and Animal Science (IAAS), Center for Crop Development and Agrobiodiversity Conservation (CCDABC), CSB, etc) across Nepal. The key actors were Nepal Genebank, LI-BIRD and Bioversity International. Through policy dialogues, research presentations and interactive discussions, decision-makers were made aware of the ecological, economic and social advantages of integrating local seed systems into formal frameworks. By highlighting the superior adaptability, resilience and sustainability of farmers' varieties. These engagements successfully influenced authorities to rethink existing policies that primarily favoured uniform and exotic varieties.

To further strengthen the case for farmers' variety registration, multiple strategies were adopted to enhance the value and understanding of local seed systems. By demonstrating that farmers' varieties contribute to food security, nutrition, climate resilience and economic benefits, policymakers were encouraged to re-evaluate seed regulations and make them more inclusive of traditional landraces. These efforts have played a significant role in shaping Nepal's evolving seed policies, ensuring that farmers' varieties gain formal recognition, legal protection and access to incentives

Seed policy reforms and legal provisions

Further progress was made with the 2022 amendment to the Seed Act, which provided legal recognition of farmers' varieties and delegated authority to provincial governments to notify and regulate local landraces. The amendment also introduced a provision for collective ownership, allowing CSBs and farmer groups to hold joint ownership of their traditional landraces. Additionally, a committee was proposed in this Act involving key agricultural officers and progressive farmers for the management and governance of local seeds, ensuring that decision-making power remained with local farming communities.

The Seed Regulation 2024 has significantly strengthened the recognition and registration of farmers' varieties by introducing detailed registration formats (Annex 5 and 6), a Truthful Labelling System, and a clear legal definition of native and local landraces (MoALD, 2024). These revisions marked a crucial step toward securing farmers' rights, ensuring their ownership, selection, conservation and distribution of landraces while promoting the sustainable use of Nepal's rich agrobiodiversity. A separate annex has been designated for native and local varieties, providing structured guidelines for

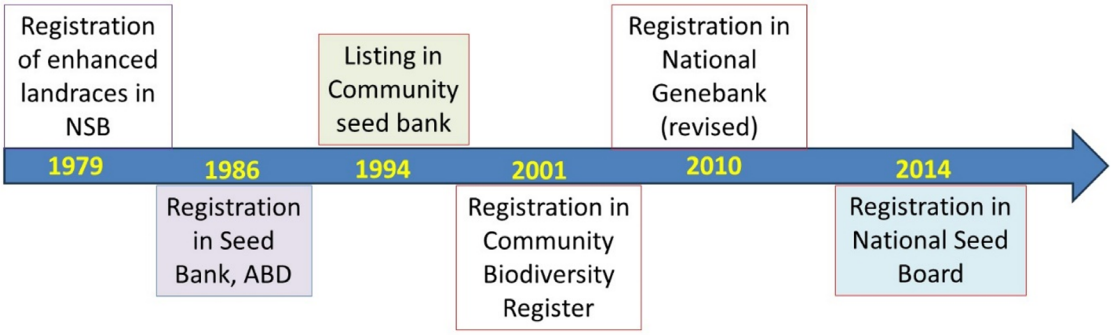


Figure 1. Informal and formal registration practices of native landraces in Nepal. NSB, National Seed Board; ABD, Agriculture Botany Division. Source: Sthapit et al, 1998; Joshi and Witcombe, 2003; Subedi et al, 2005; Subedi et al, 2013; Joshi, 2017; Joshi et al, 2017a; Thapa et al, 2019; SQCC, 2024.

their identification and registration. Native varieties are those that have been grown for generations, whereas local varieties are defined as those that have been cultivated in a particular locality for over 50 years (Joshi et al, 2020a; MoALD, 2022). As per Annex 5, farmers or communities can register their native varieties. However, the registration of local varieties follows a different approach under Annex 6, where farmers themselves cannot apply; instead, the process is managed through relevant institutions or governing bodies.

To maintain quality and traceability, farmers' varieties are classified into distinct seed classes (MoALD, 2024). For seed-propagated crops, the categories include Source Seed and Improved Seed, while for vegetatively propagated crops, the designated classes are Mother Plant and Improved Plant. Furthermore, the Truthful Labelling System ensures transparency and credibility in the marketing of farmers' varieties. It guarantees that registered varieties meet quality standards while preserving their genetic diversity and adaptability, which are essential for sustainable agriculture and climate resilience. Truthful Labelling System in the seed system refers to a regulatory approach that allows seed producers or sellers to market seeds without undergoing official certification, provided they label the seed packages accurately and honestly with all essential quality information.

Present status and ongoing initiatives

The farmers' variety registration system in Nepal has been gaining momentum, with increasing participation from farming communities. Between 2014 and 2024, farmers from 11 districts successfully registered 14 landraces of 10 different crops, excluding rice (Figure 2) (SQCC, 2024). Additionally, 13 rice landraces from 8 districts have been formally registered in the NSB (Figure 3). Encouraged by these successes, many farmers across the country are now expressing interest in registering their local varieties, recognizing the economic, nutritional and cultural benefits associated with their conservation and commercialization. Landraces are widely valued for their purity, quality, taste, nutrition, and health benefits (Gauchan et al, 2020a; Joshi et al, 2017a). Recognizing these advantages, registration has been promoted as a good conservation practice to ensure that farmers continue using, preserving and improving their traditional seed varieties. This approach aligns with the broader strategy of "conservation through use," which encourages sustainable management of agrobiodiversity (Joshi et al, 2020a; 2020c).

Several institutions are actively engaged in supporting and strengthening the registration system. Key stakeholders such as Nepal Genebank, LI-BIRD, the Department of Agriculture (DoA), SAHAS-Nepal, CSBs and community genebanks are working collaboratively on various initiatives. These organizations are facilitating capacity enhancement programmes and regular interaction sessions focusing on the importance of native landraces, the registration process, and the legal aspects of Plant Variety Protection (PVP) and Farmers' Rights (FR). Additionally, access and benefit-sharing (ABS) mechanisms for AGRs are being developed, along with the promotion of 101 good conservation practices (Joshi et al, 2020a; 2020c). For the first time in Nepal, two legal definitions have been introduced to distinguish between native (indigenous) and local landraces (MoALD, 2024). This marks a significant milestone in seed policy, as it provides a clear framework for their identification, protection and promotion. Concurrently, various scientific studies, including collection, characterization, evaluation and DNA analysis, are

Figure 2. Registered landraces other than rice crops by communities. Source: Joshi et al, 2017a; SQCC, 2024.

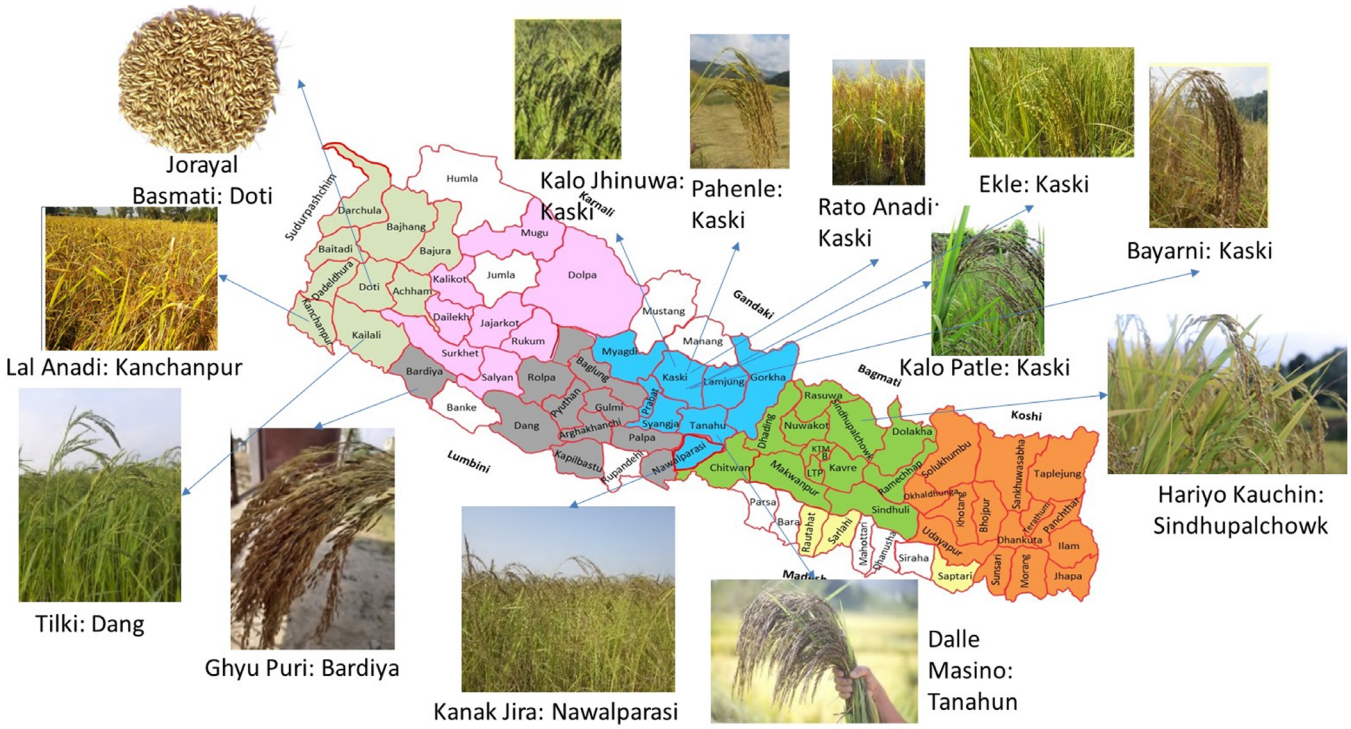


Figure 3. Registered rice landraces by communities. Source: SQCC, 2024; Joshi et al, 2017a.

being conducted to document farmers' varieties for potential geographical indication (GI) certification (Joshi et al, 2017a; Bajracharya et al, 2012; Gauchan et al, 2020a; Joshi, 2017). GI is very important for community to gain monetary benefit, which ultimately promotes native and localized landraces.

The development and enhancement of farmers' varieties is being actively pursued through participatory landrace enhancement (PLE) and participatory plant breeding (PPB), with a focus on site-specific variety development tailored to local agroecological conditions. Through PPB, farmers can develop new varieties which can be registered with support from relevant organizations. Advanced DNA fingerprinting, tissue banking, and genetic diversity assessments are also ongoing to further document and preserve Nepal's rich agrobiodiversity. Beyond registration, a variety of community-led conservation activities are regularly organized by Nepal Genebank, LI-BIRD, the Department of Agriculture and some municipalities (Joshi et al, 2020b). These include diversity blocks, diversity fairs, diversity field schools and repatriation programmes aimed at promoting traditional varieties and reintroducing lost landraces to their places of origin (Joshi et al, 2020a). Additionally, market-linkage strategies, such as the promotion of Haat Bazaars (local markets), collection centres, homestays, and product diversification initiatives, are being implemented to strengthen the connection between primary producers and primary consumers.

Impact of farmers' variety registration

The registration of farmers' varieties has had a transformative impact on Nepal's seed system, agricultural biodiversity and farming communities. One of the most significant outcomes is the conservation of site-specific varieties and genetic diversity, ensuring that traditional landraces are preserved in their natural agroecological zones. This has strengthened on-farm conservation, allowing farmers to maintain and improve their varieties while continuing traditional seed-saving practices (Joshi and Witcombe, 2003; Shrestha and Rana, 2018; Joshi et al, 2020b).

Farmers' dependency on external seed sources has been eliminated, as seed cycles are now entirely managed by farmers themselves (MoALD, 2024) for these varieties. With this shift, farmers have reclaimed their seed sovereignty, reinforcing their rights over seed production, selection and distribution. As a result, they are no longer reliant on commercial seed companies or government agencies for seed supply, ensuring greater self-sufficiency and sustainability in their farming practices (Gurung et al, 2020).

The ability to sell seeds formally has provided economic benefits to farming communities. With registered varieties, farmers can legally market their seeds, opening up new income opportunities. This has empowered smallholder farmers, many of whom previously engaged in informal seed exchanges without financial returns. Additionally, the recognition of farmers' varieties under the formal system has made them eligible for government incentives and support programmes (Gauchan et al, 2017).

Capacity building has been a crucial outcome of the registration process. Farmers have enhanced their skills in seed production, maintenance, labelling, marketing, and managing varietal diversity. This has led to better seed quality, improved market access and stronger value chains for traditional seeds. Moreover, the collaboration between farmers and plant breeders has been strengthened, fostering knowledge exchange and participatory breeding initiatives. This partnership ensures that both traditional knowledge and modern scientific approaches contribute to the development of climate-resilient, high-performing varieties (Sthapit et al, 1998; Joshi et al, 2017b).

Future directions and recommendation

Despite progress in registering farmers' varieties in Nepal, several challenges hinder their adoption and recognition in the formal seed system. Many agriculturists and policymakers believe these varieties cannot meet national demand. Their high genetic diversity, while resilient, complicates identification, monitoring and mechanization, limiting large-scale commercial use. Farmers have faced difficulties maintaining seed quality and adhering to standard practices. CSBs often struggle to market seeds, and farmers may lack the confidence or technical knowledge to defend their varieties during registration. The absence of incentives – such as financial support, price premiums or preferential treatment – along with registration costs and long-term maintenance requirements, discourages participation. The focus of national extension programmes on hybrid seeds further limits opportunities for traditional varieties. Additionally, unclear legal frameworks (for example lack of farmers' rights) around ownership rights create uncertainty about farmers' control over their genetic resources. Overcoming these barriers requires stronger policy support, financial incentives and legal clarity to ensure farmers' varieties are registered, protected, promoted and widely adopted.

The future of farmers' variety registration in Nepal should be shaped by scientific advancements, policy reforms, and farmer-centric approaches to ensure recognition, conservation and commercialization of traditional varieties. To achieve this, the registration system must evolve to become more inclusive, decentralize, and supportive of farmers' rights. One of the key reforms should be the provision of population registration for highly heterogeneous traditional mixtures, evolutionary population and component registration of mixtures as is done in inbred lines of F1 hybrids. This would allow for greater flexibility in recognizing diverse seed populations that do not fit within conventional registration criteria. Additionally, the registration process should be decentralized, allowing both local and provincial-level authorities to facilitate the process, making it more accessible to smallholder farmers.

To protect Nepal's rich seed heritage, GI rights should be established for many farmers' varieties and their value-added products (Joshi et al, 2017a). Many traditional landraces hold unique characteristics tied to specific regions, making them eligible for GI branding and marketing, which could enhance their economic value (Joshi, 2017; Shrestha and Rana, 2018). Furthermore, the PVP and FR framework should be adopted nationwide to safeguard farmers' traditional knowledge, breeding efforts and seed sovereignty. Establishing a clear ABS mechanism for AGRs will also be crucial in ensuring that farmers receive fair compensation for their contributions to biodiversity conservation and seed development (Gauchan et al, 2018).

The enhancement and conservation of landraces should be approached through site-specific breeding and value addition, considering the unique agroecological and household-specific requirements of farmers (Joshi et al, 2020c; 2023a). Many traditional varieties have proven to be superior in ecological yield and food health index (FHI) compared to modern improved varieties. Since farmers are well aware of the taste, nutritional benefits, and medicinal properties of their varieties, integrating FHI-based evaluation into mainstream seed selection will further strengthen the promotion and utilization of farmers' varieties. Another critical area of focus should be seed storage. Modern agriculture has shifted towards plastic-based seed storage, which may not always be environmentally sustainable or suitable for long-term seed viability. Traditional nature-positive storage methods, such as those made from clay, wood, bamboo, leaves, bark and fruit-based materials, should be encouraged, as they align better with farmers' seasonal storage practices while supporting environmental sustainability.

A localized seed system with a global market approach should be adopted to ensure farmers maintain control over their seeds while expanding opportunities for commercialization (Joshi et al, 2020a). Farmers should be allowed to market native and local varieties without mandatory registration, as these varieties should automatically qualify for government incentives, legal benefits, and services. Furthermore, landraces should be recognized as private goods rather than public goods, since individual households have preserved unique seed lines over generations.

The integration of advanced scientific tools such as phenomics, genomics and foodomics will enhance the study and improvement of farmers' varieties. Gene tagging and mapping should be applied to document and utilize key genetic traits, ensuring the conservation and sustainable use of these valuable genetic resources. Additionally, a searchable online database should be developed to systematically document Nepal's registered and native farmers' varieties, making information more accessible to farmers, researchers and policymakers.

Lastly, the scope of variety registration should be expanded to include other crops such as fruits, forages, medicinal plants and other AGRs. Many of these crops hold significant economic, medicinal and cultural value, making their conservation and formal recognition equally important. So far, registration of farmers' varieties has been only on orthodox seed, and due attention should also be given to non-orthodox crops (Pokhrel et al, 2017; SQCC, 2024). By adopting a multi-dimensional approach that integrates traditional knowledge with modern science, Nepal can create a more inclusive, farmer-friendly and scientifically robust farmers' variety registration system.

Conclusion

The registration of farmers' varieties in Nepal marks a significant step towards recognizing and formalizing traditional seed systems. Historically, the DUS system has been the foundation for variety registration, primarily to simplify identification and monitoring. However, this approach has often excluded farmers' varieties, which naturally exhibit higher genetic diversity and adaptability. Nepal's formal seed system gradually separated seed production from farming communities, leading to increased dependency on external entities for seed supply. This detachment marginalized farmers' traditional seed systems, preventing their varieties from entering the formal market and making them ineligible for incentives, benefits and government support. However, with the introduction of farmers' variety registration, a shift has occurred – allowing farming communities to formally register, produce and market their own seeds, thereby re-establishing their role in the seed production cycle. This has also led to the legal recognition of genetic diversity within traditional varieties, promoting agrobiodiversity conservation through policy support.

Despite this progress, challenges remain. The current registration system still struggles to accommodate highly

heterogeneous varieties and mixtures of multiple landraces. To address this, a policy shift is necessary – where native and local varieties should be automatically eligible for formal marketing, incentives and other legal benefits without requiring extensive bureaucratic approval. Moving forward, the formal seed system should focus more on value-added products, rather than strictly controlling seed production. By doing so, Nepal can adopt a localized seed system with a globalized product approach, ensuring that farmers maintain seed sovereignty while expanding market opportunities. A balanced approach that integrates both formal and informal seed systems will be key to sustainable agriculture, biodiversity conservation and farmers' economic empowerment in Nepal.

Acknowledgements

We express our sincere gratitude to the Nepal Agricultural Research Council (NARC), Nepal Genebank, Seed Quality Control Centre (SQCC), Local Initiatives for Biodiversity, Research, and Development (LI-BIRD), Bioversity International, and SAHAS Nepal for their invaluable support and collaboration in organizing various programmes and action research initiatives. We extend our heartfelt appreciation to the Community Seed Banks (CSBs), the Community Seed Bank Association of Nepal, and the dedicated farmers who have actively contributed to the registration of their landraces, the maintenance of crop biodiversity, and their enthusiastic participation in the preparation of guidelines, training session, and participatory on-farm research.

Special acknowledgment goes to the farmers and the CSB in Dalchoki for their commendable efforts and commitment in generating critical evidence through on-farm trials, which have significantly contributed to the success of our research. We also wish to convey our deep gratitude to Madan Raj Bhatta (NAGRC), Madan Thapa (SQCC), Devendra Gauchan, and Mohan Hamal for fostering a supportive environment and taking meaningful action in advancing our shared goals.

The following projects provided financial support for this work. 1. Strengthening the scientific basis of in-situ conservation of agrobiodiversity on farm in Nepal, 1997 (in-situ global project funded by IPGRI), 2. Genetic Resources Policy Initiative (GRPI), 2003 and 2007, funded by IPGRI, 3. Strengthening National Capacities to Implement the International Treaty on Plant Genetic Resources for Food and Agriculture, 2012 (GRPI-2), funded by Bioversity International, 4. Improving seed systems for smallholder farmers' food security, i: 2013 to 2016, and ii. 2017-2021 (Seed project), 119HQ177, funded by SDC via Bioversity International, 5. Integrating traditional crop genetic diversity into technology: Using a biodiversity portfolio approach to buffer against unpredictable environmental change in the Nepal Himalayas, 2014-2020, funded by UNEP-GEF through Bioversity International, 6. Use of genetic diversity and evolutionary plant breeding for enhanced farmer resilience to climate change, sustainable crop productivity, and nutrition under rainfed conditions, 2019-2022 (EPB), A1341 funded by IFAD through Bioversity International.

Author contributions

BK Joshi: conceptualization, data collections, field experiments, field visits, interactions, writing – original, research, methodology, and editing. P Thapa: literature review, secondary data collection, focus group discussion, editing and updating information. B Prasai: literature review, policy analysis, editing and updating information. DR Bhandari: documentation of history, gap analysis, policy analysis, editing and updating information.

Conflict of interest statement

The authors have declared that no competing interests exist.

Ethics statement

Stakeholder engagement was conducted respectfully and inclusively, and all data were used solely for research purposes in line with relevant institutional and national ethical guidelines.

References

Bajracharya, J., Brown, A. H. D., Joshi, B. K., Panday, D., Baniya, B. K., Sthapit, B. R., and Jarvis, D. I. (2012). Traditional seed management and genetic diversity in barley varieties in high-hill agro-ecosystems of Nepal. *Genetic Resources and Crop Evolution*, 59(3), 389–398. https://doi.org/10.1007/s10722-011-9689-2

Bhandari, B., Joshi, B. K., Shrestha, P., Sthapit, S., Chaudhary, P., and Acharya, A. K. (2017). Custodian farmers, agrobiodiversity-rich areas, and agrobiodiversity conservation initiatives at grassroots levels in Nepal. In B. K. Joshi, H. B. K. C., and A. K. Acharya (Eds.), Conservation and utilization of agricultural plant genetic resources in Nepal: Proceedings of the 2nd National Workshop, 22-23 May 2017, Dhulikhel (pp. 92–101). NAGRC, FDD, DoA, and MoAD. https://www.researchgate.net/publication/348049968_Conservation_and_Utilization_of_Agricultural_Plant_Genetic_Resources_in_Nepal_Proceedings_of_2nd_National_Workshop

Chaudhary, P., Joshi, B. K., Thapa, K., Devkota, R., Ghimire, K. H., Khadka, K., Upadhya, D., and Vernooy, R. (2016). Interdependence on plant genetic resources in light of climate change. In B. K. Joshi, P. Chaudhary, D. Upadhya, and R. Vernooy (Eds.), Implementing the International Treaty on Plant Genetic Resources for Food and Agriculture in Nepal: Achievements and challenges (pp. 65–80). LIBIRD, NARC, MoAD, and Bioversity International. https://hdl.handle.net/10568/78421

De Jonge, B., Dey, B., and Visser, B. (2025). Developing a registration system for farmers' varieties. *Agricultural Systems*, 222, 104183. https://doi.org/10.1016/j.agsy.2024.104183

Gauchan, D., Joshi, B. K., and Bhandari, B. (2018). Farmers' rights and access and benefit-sharing mechanisms in community seed banks in Nepal. In B. K. Joshi, P. Shrestha, D. Gauchan, and R. Vernooy (Eds.), Community seed banks in Nepal: 2nd National Workshop Proceedings, 3-5 May 2018, Kathmandu, Nepal (pp. 117–132). NAGRC, LI-BIRD, and Bioversity International. https://hdl.handle.net/10568/99141

Gauchan, D., Joshi, B. K., Bhandari, B., Ghimire, K. H., Pant, S., Gurung, R., Pudasaini, N., Paneru, P. B., Mishra, K. K., and Jarvis, D. I. (2020a). Value chain development and

mainstreaming of traditional crops for nutrition-sensitive agriculture in Nepal. In D. Gauchan, B. K. Joshi, B. Bhandari, H. K. Manandhar, and D. Jarvis (Eds.), Traditional crop biodiversity for mountain food and nutrition security in Nepal: Tools and research results of the UNEP GEF Local Crop Project, Nepal (pp. 174–182). NAGRC, LI-BIRD, and the Alliance of Bioversity International and CIAT. https://himalayancrops.org/project/traditional-crop-biodiversity-for-mountain-food-and-nutrition-security-in-nepal/

Gauchan, D., Joshi, B. K., Sthapit, S., and Jarvis, D. (2020b). Traditional crops for household food security and factors associated with on-farm crop diversity in the mountains of Nepal. *The Journal of Agriculture and Environment*, 21, 31–43. https://hdl.handle.net/10568/109041

Gauchan, D., Joshi, B. K., Sthapit, S., Ghimire, K., Gautam, S., Poudel, K., Sapkota, S., Neupane, S., Sthapit, B., and Vernooy, R. (2016). Post-disaster revival of the local seed system and climate change adaptation: A case study of earthquake-affected mountain regions of Nepal. *Indian Journal of Plant Genetic Resources*, 29(3), 119–119.

Gauchan, D., Tiwari, S. B., Acharya, A. K., Pandey, K. R., and Joshi, B. K. (2017). National and international policies and incentives for agrobiodiversity conservation and use in Nepal. In B. K. Joshi, H. B. K. C., and A. K. Acharya (Eds.), Conservation and utilization of agricultural plant genetic resources in Nepal: Proceedings of the 2nd National Workshop, 22-23 May 2017, Dhulikhel (pp. 176–183). NAGRC, FDD, DoA, and MoAD. https://www.researchgate.net/publication/348049968_Conservation_and_Utilization_of_Agricultural_Plant_Genetic_Resources_in_Nepal_Proceedings_of_2nd_National_Workshop

Gurung, R., Pudasaini, N., Sthapit, S., Palikhey, E., and Gauchan, D. (2020). Seed system of traditional crops in the mountains of Nepal. In D. Gauchan, B. K. Joshi, B. Bhandari, H. K. Manandhar, and D. I. Jarvis (Eds.), Traditional crop biodiversity for mountain food and nutrition security in Nepal: Tools and research results of UNEP GEF Local Crop Project, Nepal (pp. 108–115). NAGRC, LI-BIRD, and the Alliance of Bioversity International and CIAT. http://himalayancrops.org/wp-content/uploads/2020/06/Traditional-Crop-Biodiversity-for-Mountain-Food-and-Nutrition-Security-in-Nepal.pdf

Gurung, R., Dhakal, R., Pudasaini, N., Paneru, P.B., Pant, S., Adhikari, A.R., Gautam, S., Yadav, R.K., Ghimire, K.H., Joshi, B.K., Gauchan, D., Shrestha, S., and Jarvis, D.I. (2019). Catalog of Traditional Crop Landraces of Mountain Agriculture in Nepal. LI-BIRD, Pokhara; NARC, Kathmandu and Bioversity International, Nepal. http://himalayancrops.org/project/catalogue-of-traditional-mountain-crop-landraces-in-nepal/

Joshi, B.K., Gurung, S.B., Mahat, P.M., Bhandari, B. and Gauchan, D. (2018). Intra-Varietal Diversity in Landrace and Modern Variety of Rice and Buckwheat. T*he Journal of Agriculture and Development*, 19, 1-8. https://hdl.handle.net/10568/97576

Joshi, B. K. (2017). Plant breeding in Nepal: Past, present, and future. Journal of Agriculture and Forestry University, 1, 1–33. http://afu.edu.np/sites/default/files/Plant_breeding_in_Nepal_Past_Present_and_Future_BK_Joshi.pdf

Joshi, B. K., Acharya, A. K., Gauchan, D., Singh, D., Ghimire, K. H., and Sthapit, B. R. (2017a). Geographical indication: A tool for supporting on-farm conservation of crop landraces and for rural development. In B. K. Joshi, H. B. K. C., and A. K. Acharya (Eds.), Conservation and utilization of agricultural plant genetic resources in Nepal: Proceedings of the 2nd National Workshop, 22-23 May 2017, Dhulikhel (pp. 50–62). NAGRC, FDD, DoA, and MoAD. https://www.researchgate.net/publication/348049968_Conservation_and_Utilization_of_Agricultural_Plant_Genetic_Resources_in_Nepal_Proceedings_of_2nd_National_Workshop

Joshi, B. K., Bhatta, M. R., Ghimire, K. H., Khanal, M., Gurung, S. B., Dhakal, R., and Sthapit, B. R. (2017b). Released and promising crop varieties of mountain agriculture in Nepal (1959-2016). LI-BIRD, Pokhara; NARC, Kathmandu; and Bioversity International. https://hdl.handle.net/10568/80892

Joshi, B. K., Gauchan, D., Bhandari, B., and Jarvis, D. (Eds.). (2020a). Good practices for agrobiodiversity management. NAGRC, LI-BIRD, and Alliance of Bioversity International and CIAT. https://hdl.handle.net/10568/109625

Joshi, B. K., Ghimire, K. H., Neupane, S. P., Gauchan, D., and Mengistu, D. K. (2023a). Approaches and advantages of increased crop genetic diversity in the fields. Diversity, 15(603). https://doi.org/10.3390/d15050603

Joshi, B. K., Gorkhali, N. A., Pradhan, N., Ghimire, K. H., Gotame, T. P., K. C., P., Mainali, R. P., Karkee, A., and Paneru, R. B. (2020b). Agrobiodiversity and its conservation in Nepal. *Journal of Nepal Agricultural Research Council*, 6, 14–33. https://doi.org/10.3126/jnarc.v6i0.28111

Joshi, B. K., Humagain, R., Dhakal, L. K., and Gauchan, D. (2020c). Integrated approach of national seed systems for assuring improved seeds to smallholder farmers in Nepal. In R. B. Shrestha, M. E. Penunia, and M. Asim (Eds.), Strengthening seed systems – Promoting community-based seed systems for biodiversity conservation and food and nutrition security in South Asia (pp. 181–194). SAARC Agriculture Center, Bangladesh; Asian Farmers' Association, the Philippines; and Pakistan Agricultural Research Council, Pakistan. https://www.researchgate.net/publication/342491585_Integrated_Approach_of_National_Seed_Systems_for_Assuring_Improved_Seeds_to_the_Smallholder_Farmers_in_Nepal

Joshi, B.K., Ojha, P., Gauchan, D., Ghimire, K.H., Bhandari, B. and KC, H.B. (2020d). Nutritionally unique native crop landraces from mountain Nepal for geographical indication right. In D. Gauchan, B.K. Joshi, B. Bhandari, H.K. Manandhar and D. Jarvis (Eds.), Traditional Crop Biodiversity for Mountain Food and Nutrition Security in Nepal . Tools and Research Results of the UNEP GEF Local Crop Project, Nepal (pp. 87-99). NAGRC, LI-BIRD and the Alliance of Bioversity International and CIAT; Kathmandu, Nepal. https://himalayancrops.org/project/traditional-crop-biodiversity-for-mountain-food-and-nutrition-security-in-nepal/

Joshi, B.K., Gauchan, D., Ghimire, K.H. and Ayer, D.K. (2022). Five-cell Analysis (FCA). In B.K. Joshi, D. Gauchan and .D.K Ayer (Eds.), Participatory agrobiodiversity tools and methodologies (PATaM) in Nepal, (pp. 24-25). NAGRC, LI-BIRD, and Alliance of Bioversity International and CIAT; Kathmandu, Nepal. https://api.giwms.gov.np/storage/75/posts/1685027635_2.pdf

Joshi, K. D., and Witcombe, J. (2003). The impact of participatory plant breeding (PPB) on landrace diversity: A case study for high-altitude rice in Nepal. *Euphytica*, 134, 117–125. doi: https://doi.org/10.1023/A:1026151017274

Karkee, A., Magar, P. B., Thapa, P., Mainali, R. P., Ghimire, K. H., Joshi, B. K., and Bhattarai, M. (2023). Evaluation of barley landraces for yellow rust disease resistance along with grain yield. In Proceedings of the 32nd National Winter Crops Workshop, 31 May–2 June 2022, Lumle, NARC (pp. 429–439). https://www.researchgate.net/publication/367074827

LI-BIRD and The Development Fund. (2017). Farmers' seed systems in Nepal: Review of national legislations. Pokhara, Nepal. https://libird.org/farmers-seed-system-in-nepal-review-of-national-legislations/

Ministry of Agricultural Development (MoAD). (2013). Seed Regulation-2013. Ministry of Agricultural Development, Government of Nepal, Kathmandu, Nepal. (In Nepali).

Ministry of Agriculture (MoA). (1997). The Seeds Regulation-1997. Ministry of Agriculture, His Majesty's Government, Nepal. (In Nepali).

Ministry of Agriculture and Livestock Development (MoALD). (2022). Seed Act-1988 (2nd amendment). Ministry of Agriculture and Livestock Development, Government of Nepal, Kathmandu, Nepal. (In Nepali).

Ministry of Agriculture and Livestock Development (MoALD). (2024). Seed Regulation. Ministry of Agriculture and Livestock Development, Government of Nepal, Kathmandu, Nepal. (In Nepali).

NAGRC. 2024. Annual Report 2080/81 (2023/24). National Agriculture Genetic Resources Centre, NARC, Khumaltar, Lalitpur, Nepal.

Neupane, S.P., Joshi, B.K., Aair, D., Ghimire, K.H., Gauchan, D., Karkee, A., Jarvis, D.I., Mengistu, D.K., Grando, S., and Ceccarell, S. (2023). Farmers' preferences and agronomic evaluation of dynamic mixtures of rice and bean in Nepal. *Diversity*, 15, 660. https://doi.org/10.3390/d15050660

Pokhrel, S., Joshi, B. K., Gauchan, D., and Acharya, A. K. (2017). Utilization of agricultural plant genetic resources in research, breeding, production, nutrition, and distribution. In B. K. Joshi, H. B. K. C., and A. K. Acharya (Eds.), Conservation and utilization of agricultural plant genetic resources in Nepal: Proceedings of the 2nd National Workshop, 22-23 May 2017, Dhulikhel (pp. 39–49). NAGRC, FDD, DoA, and MoAD, Kathmandu, Nepal. https://www.researchgate.net/publication/348049968_Conservation_and_Utilization_of_Agricultural_Plant_Genetic_Resources_in_Nepal_Proceedings_of_2nd_National_Workshop

Poudel, D., Sthapit, B., and Shrestha, P. (2015). An analysis of social seed network and its contribution to on-farm conservation of crop genetic diversity in Nepal. *International Journal of Biodiversity*, 2015, Article ID 312621. https://doi.org/10.1155/2015/312621

Shrestha, P., and Rana, R. B. (2018). Community seed banks in Nepal: Safeguarding agricultural biodiversity and strengthening local seed systems. In B. K. Joshi, P. Shrestha, D. Gauchan, and R. Vernooy (Eds.), Community seed banks in Nepal: 2nd National Workshop proceedings, 3–5 May 2018, Kathmandu, Nepal (pp. 45–67). NAGRC, LI-BIRD, and Bioversity International. https://hdl.handle.net/10568/99141

Shrestha, P., Sthapit, S., and Paudel, I. (2013). Community seed banks: A local solution to increase access to quality and diversity of seeds. In P. Shrestha, R. Vernooy, and P. Chaudhary (Eds.), Community seed banks in Nepal: Past, present, future. Proceedings of a national workshop (pp. 61–75). LI-BIRD/USC Canada Asia/Oxfam/The Development Fund/IFAD/Bioversity International. https://hdl.handle.net/10568/68933

SQCC. 2024. Notified or denotified varieties updated 2081-07-23. Seed Quality Control Center, MoALD, Kathmanud, Nepal (in Nepali language).

Sthapit, B. (2013). Emerging theory and practice: Community seed banks, seed system resilience, and food security. In P. Shrestha, R. Vernooy, and P. Chaudhary (Eds.), Community seed banks in Nepal: Past, present, future. Proceedings of a national workshop (pp. 16–40). LI-BIRD/USC Canada Asia/Oxfam/The Development Fund/IFAD/Bioversity International. https://hdl.handle.net/10568/68933

Sthapit, B., and Jarvis, D. (2003). Implementation of on-farm conservation in Nepal. In D. Gauchan, B. R. Sthapit, and D. I. Jarvis (Eds.), Agrobiodiversity conservation on-farm: Nepal's contribution to a scientific basis for national policy recommendations (pp. 8–12). IPGRI. https://hdl.handle.net/10568/105239

Sthapit, B., Gauchan, D., Sthapit, S., Ghimire, K. H., Joshi, B. K., De Santis, P., and Jarvis, D. (2019). Sourcing and deploying new crop varieties in mountain production systems. In O. T. Westengen and T. Winge (Eds.), Farmers and plant breeding: Current approaches and perspectives (pp. 196–216). Routledge. https://hdl.handle.net/10568/105452

Sthapit, B., Joshi, K., Rana, R., and Subedi, A. (1998). Spread of varieties from participatory plant breeding in the high altitude villages of Nepal. LI-BIRD Technical Report Series. LI-BIRD. https://libird.org/spread-of-varieties-from-participatory-plant-breeding-in-high-altitude-villages-of-nepal/

Subedi, A., Devkota, R., Paudel, I. P., and Subedi, S. (2013). Community biodiversity registers in Nepal: Enhancing the capabilities of communities to document, monitor, and take control over their genetic resources. In W. S. de Boef, N. Peroni, A. Subedi, M. H. Thijssen, and E. O'Keeffe (Eds.), Community biodiversity management: Promoting resilience and the conservation of plant genetic resources (pp. 177–188). Earthscan and Bioversity International. https://doi.org/10.4324/9780203130599

Subedi, A., Sthapit, B., Shrestha, P., Gauchan, D., and Upadhyay, M. (2005). Emerging methodology of community biodiversity register: A synthesis. In A. Subedi, B. R. Sthapit, M. P. Upadhyay, and D. Gauchan (Eds.), Learning from community biodiversity register in Nepal: Proceedings of the national workshop, 27–28 October 2005, Khumaltar, Nepal (pp. 75–83). https://idl-bnc-idrc.dspacedirect.org/bitstreams/4cbf2874-f54e-4307-96f7-cd52393ad51d/download

Thapa, M., Sapkota, D., and Joshi, B. K. (2019). Crop groups based on national list: Released, registered, denotified, formal, and informal. In B. K. Joshi and R. Shrestha (Eds.), Proceedings of the national workshop of the working groups of agricultural plant genetic resources (APGRs), 21–22 June 2018 (pp. 36–44). NAGRC, Kathmandu, Nepal. https://www.researchgate.net/publication/334400726_crop_group_based_on_national_list_released_registered_denotified_formal_and_informal

Upreti, B. R., and Upreti, Y. G. (2002). Factors leading to agro-biodiversity loss in developing countries: The case of Nepal. *Biodiversity and Conservation*, 11(8), 1607–1621. Doi: https://doi.org/10.1023/A:1016862200156

Zhu, Y., Chen, H., Fan, J., Wang, Y., Li, Q., Chen, J., Fan, J., Yang, S., Hu, L., Leunak, H., Mewk, T.W., Tengk, P.S., Wanak, Z. and Mundt, C.C. (2000). Genetic diversity and disease control in Rice. *Nature*, 406, 718-722.

Genetic Resources

ORIGINAL ARTICLE

Genetic Resources (2026), 7 (13), 77–88
DOI: 10.46265/genresj.KPMC9702
https://www.genresj.org
ISSN: 2708-3764

Morphological variation of *Pseudocedrela kotschyi* in Benin: zonal patterns and conservation insights

Tonankpon Aymar Guy Deguenonvo[a], Dowo Michée Adjacou[a,*], Rodrigue Idohou[b,c], Reine Sodedja[a], Florent Eudes Dagbédji Sobakin[a], Thierry Dehouegnon Houehanou[a], Gérard Nounagnon Gouwakinnou[a], Armand Kuyema Natta[d], Frank Hellwig[e]

[a] *Research Unit of Biodiversity Conservation at the Interface People, land use and Climate changes (UR-BIPLaC), Laboratory of Ecology, Botany and Plant Biology (LEB), Faculty of Agronomy, University of Parakou, 03 BP: 125, Parakou, Benin*
[b] *Ecole de Gestion et de Production Végétale et Semencière, Université Nationale d'Agriculture, BP 43, Kétou, Benin*
[c] *Laboratoire de Biomathématiques et d'Estimations Forestières, Faculté des Sciences Agronomiques, Université d'Abomey-Calavi, 04 BP 1525, Cotonou, Benin*
[d] *Laboratory of Ecology, Botany and Plant Biology (LEB), Faculty of Agronomy, University of Parakou. 03 BP: 125, Parakou, Benin*
[e] *Department for Systematic Botany with Hausknecht Herbarium and Botanical Garden, Friedrich-Schiller-University Jena, Philosophenweg 16, 07743 Jena, Germany*

Abstract: *Pseudocedrela kotschyi* is a socio-economically important species for rural communities in sub-Saharan Africa. This study aims to identify the environmental drivers shaping the morphological traits of the species across different biogeographical zones for guiding effective conservation and domestication strategies. Measurements were taken from 3,086 fruits and 2,851 leaves that were collected in these zones. Principal Component Analysis (PCA) and hierarchical classification identified distinct morphological groups, while Multiple Correspondence Analysis (MCA) assessed the relationship between morphotypes and biogeographical zones. Morphological traits varied significantly between zones ($p < 0.001$). The Guineo-Congolian zone had the longest (10.27 ± 0.05cm) and heaviest fruits (51.59 ± 0.39g), while the Sudano-Guinean zone had the heaviest seeds (0.69 ± 0.01g). Three morphotypes were identified: morphotype 1 had small fruits with light seeds; morphotype 2 had long fruits with numerous seeds whereas morphotype 3 had heavy fruits with large seeds. The distortion features differ from one morphotype to another. Although certain traits were influenced by temperature and precipitation, relationships between morphology and climate remained weak. These findings highlight the importance of conservation strategies that are adapted to regional specificities and local environmental pressures. Tailoring conservation and domestication strategies to the distinct morphotypes and their associated ecological zones could enhance the sustainable use of, and resilience to climate change pressures experienced by *Pseudocedrela kotschyi*.

Keywords: Biogeographical zones, fruit morphotypes, climatic influence, domestication potential, morphological variability

Citation: Deguenonvo, T. A. G., Adjacou, D. M., Idohou, R., Sodedja, R., Sobakin, F. E. D., Houehanou, T. D., Gouwakinnou, G. N., Natta, A. K. and Hellwig, F. (2026) "Morphological variation of *Pseudocedrela kotschyi* in Benin: zonal patterns and conservation insights", *Genetic Resources*, 7(13), pp. 77–88. doi: 10.46265/genresj.KPMC9702.

Introduction

In West Africa, over recent years, vegetation cover has undergone continuous and severe disturbances, particularly within natural formations (Dossa *et al*, 2021). These formations are experiencing unprecedented degradation due to intense anthropogenic pressure (Ouattara *et al*, 2022) and climate variability, which negatively impact biodiversity (Zida *et al*, 2020). Furthermore, it is estimated that 20–35% of the African tropical flora is potentially threatened with extinction (Stévart *et al*, 2019). This loss is having consequences for the rich and diverse forest ecosystems, which play a crucial role in regulating greenhouse gases, maintaining the climate balance and meeting various needs of rural populations (Ouattara *et al*, 2022). In Benin, the most threatened valuable timber species include *Pterocarpus erinaceus* Poir., *Afzelia africana* Sm. Ex Pers, *Khaya senegalensis* (Desr.) A.Juss.,

* Corresponding author: Dowo Michée Adjacou (micheadjacou94206471@gmail.com)

Received: 04.11.2025 | Accepted: 15.01.2026 | Published online: 25.02.2026

Prunus africana (Hook.f.) Kalkman, *Anogeissus leiocarpa* (DC.) Guill. et Perr., and *Pseudocedrela kotschyi* (Schweinf.) Harms (Yaoitcha *et al*, 2016).

The genus *Pseudocedrela* is one of the three endemic genera (*Pseudocedrela, Vitellaria* and *Haemastostaphis)* in the Sudanian Regional Centre of Endemism (SRCE) (White, 1983). It contains a single species, *Pseudocedrela kotschyi* (Schweinf.), a forest tree which is exploited similarly to other high-value species but which faces regeneration challenges (Deguenonvo *et al*, 2020) due to bushfires and rodent predation of seeds. In addition, the seeds are highly susceptible to insect attacks (Grubben, 2008) and must be sown immediately after harvest.

Given these persistent threats, effective and sustainable management strategies to conserve this species are required. Such strategies depend on the morphological variability of its traits, which remains poorly documented. For instance, fruit size analysis can help identify vigorous shrubs producing large fruits and robust seeds suitable for agroforestry (Daï *et al*, 2024). This knowledge also supports the preservation of underutilized morphotypes that are valuable for maintaining future genetic diversity (Houehanou *et al*, 2019; 2023). This aspect is particularly critical for a species widely distributed across Benin's distinct biogeographical zones, namely, the Sudanian, Sudano-Guinean, and Guineo-Congolian zones. This is the case for *P. kotschyi*, for which studies of morphological characteristics across environmental gradients remain incomplete. Studies of African tropical species have demonstrated that morphological variability is essential for understanding their responses to climatic gradients (Avakoudjo *et al*, 2021; Hounkpèvi *et al*, 2020; Konda *et al*, 2025).

Morphological traits are fundamental not only for species identification but also for understanding their ecological adaptation and potential resilience to environmental changes. These traits are frequently employed to differentiate between taxa, particularly in species with broad geographic ranges or those exposed to variable environmental conditions. For example, Konda *et al* (2025) showed that morphological traits in *Pterocarpus erinaceus* Poir. contribute to its adaptability and resilience, allowing the species to thrive in diverse environments. Other studies on *Mangifera indica* L. have revealed substantial morphological variability among local varieties, particularly in fruit traits (e.g. fruit shape, fruit length, length of stone fibre), which can be used to classify and differentiate regional types (Adjacou *et al*, 2022; 2024; Yusuf *et al*, 2020). These studies emphasize the importance of morphological traits as reliable indicators for assessing species diversity and environmental adaptation, thereby providing a robust foundation for conservation and domestication efforts. This variability underscores the significance of morphological evaluations in agricultural biodiversity and conservation strategies. One illustrative case is the study by Ikabanga *et al* (2017) whose classification has remained controversial for over a century. Studies combining chloroplast and nuclear DNA sequences show the existence of several phylogenetic clades in this taxon, with some occurring in sympatry in western Central Africa suggesting the existence of at least two species. By combining genetic and morphological markers, we aim to assess the species delimitation in the Santiria species complex. Morphological trait (trunk, leaflet, flower and fruit characteristics, which highlighted the role of morphological traits in distinguishing species within the African tree genus *Santiria* (Burseraceae).

P. kotschyi is a species that is widely distributed across Sudanian and Sudano-Guinean zones, with an irregular distribution, locally common and gregarious (Arbonnier, 2019). It is also found in areas prone to flooding (Diarra *et al*, 2016). The most effective propagation techniques for *P. kotschyi* are direct seeding and vegetative propagation (Deguenonvo *et al*, 2020). However, the seeds must be sourced from natural populations and be properly stored after drying. Regarding vegetative propagation, only root cuttings have shown a satisfactory regeneration rate (Deguenonvo *et al*, 2020). Nevertheless, the root cuttings must be obtained from trees with good morphological traits, particularly with regard to diameter, since the size of the cuttings significantly influences their regeneration rate (Deguenonvo *et al*, 2020). *P. kotschyi* is of great importance to local communities in Benin due to its multiple uses: food, medicinal, industrial and technological (Deguenonvo *et al*, 2023b). Previous studies have investigated its propagation potential through seeds, stem cuttings, and root cuttings (Deguenonvo *et al*, 2020), as well as the structural and ecological characteristics of its populations (Moussilimi *et al*, 2022), and the synergy between climate dynamics, species distribution and structural parameters (Deguenonvo *et al*, 2024). Additionally, several studies have explored various ecological and anthropogenic factors affecting *P. kotschyi*. These include the influence of biogeographical zones and anthropogenic disturbances on its floristic composition and habitat diversity (Deguenonvo *et al*, 2023a), the relationship between stand structure and population dynamics (Assédé *et al*, 2012), and the uses, cultural significance, and fire-related threats to the species (Deguenonvo *et al*, 2023b). Recognized as a key species in tropical and subtropical regions of Africa (Alhassan *et al*, 2021), *P. kotschyi* is one of the most exploited species in Benin. Its populations are facing challenges related to anthropogenic disturbances (Deguenonvo *et al*, 2023a) and the effects of climate change (Deguenonvo *et al*, 2024).

Morphological characterization is essential in the domestication processes of indigenous species. One of the key steps in the characterization of morphotypes is identifying the most discriminating morphological descriptors. Despite existing morphological research on native species in Benin, none has specifically addressed architectural, habit-related and productivity-related discriminating descriptors or morphotypes of *P. kotschyi* in relation to climatic gradients. Yet, knowledge of such descriptors is crucial for plant breeding and varietal selection programs. Moreover, it enables the identification of relevant morphological descriptors and those associated with climatic and environmental factors. In general, plant species variability is expressed in both vegetative and reproductive traits (Mars and Marrakchi, 2000).

Given the growing interest in promoting the cultivation of *P. kotschyi* in Benin, assessing the species morphotype potential is essential. In this context, identifying the extent of morphological trait variability and the most distinctive morphotypes is a crucial step toward supporting domestication and selection programmes. This study, therefore, aims to assess the morphological trait variability of *P. kotschyi* across biogeographical zones in Benin and to examine the climatic and environmental factors driving this variation. Specifically, the objectives are to: (1) evaluate the morphological characteristics of *P. kotschyi* fruits and

leaves across Benin's biogeographical zones, (2) identify fruit morphotypes that could be used in selection programs across these zones, and (3) analyze the influence of climatic gradients and environmental variables on fruit characteristics of *P. kotschyi*.

Materials and methods

Study area

This study was conducted in Benin (6°30' – 12°30'N, 1° – 3°40'E; ~ 112,622km²), located in West Africa, which is characterized by three distinct biogeographical zones: the Guineo-Congolian zone the Sudano-Guinean zone and the Sudanian zone (Akoègninou *et al*, 2006) (Figure 1). The Guineo-Congolian zone features a subequatorial climate with bimodal rainfall, experiencing a major rainy season from April to July and a high humidity range of 85% to 90% (Adomou *et al*, 2006). Temperatures in this zone range between 23°C and 32°C, with annual precipitation varying from 950mm to 1,400mm and are characterized by semi-deciduous forests on ferralitic soils. The Sudano-Guinean zone, located between 7° and 8°30'N, exhibits complex precipitation patterns influenced by both southern and northern climatic systems, with an average annual rainfall of 1,200mm, a temperature of 27°C, and 60% humidity. This zone is mainly characterized by wooded savannas and open forests (Adomou *et al*, 2006). The Sudanian zone, located between 8°30' and 12°30'N, has a unimodal rainfall pattern, with annual precipitation ranging from 900mm to 1,100mm and an average temperature of 27.5°C (Adomou *et al*, 2006).

Sampling and data collection

Three forests were considered in the Sudanian zone, four in the Sudano-Guinean zone, and only one in the Guineo-Congolian zone. Across these three biogeographical zones, a total of 150 randomly selected individuals belonging to 13 populations of *P. kotschyi* were identified: 59 were found in the Guineo-Congolian zone, 54 in the Sudano-Guinean zone and 37 in the Sudanian zone. All samples were georeferenced using a GPS (see Table 1). Samples from Dogo-Kétou, situated within the Guineo-Congolian–Sudano-Guinean transition zone, were classified in the Guineo-Congolian zone for analytical purposes. This decision

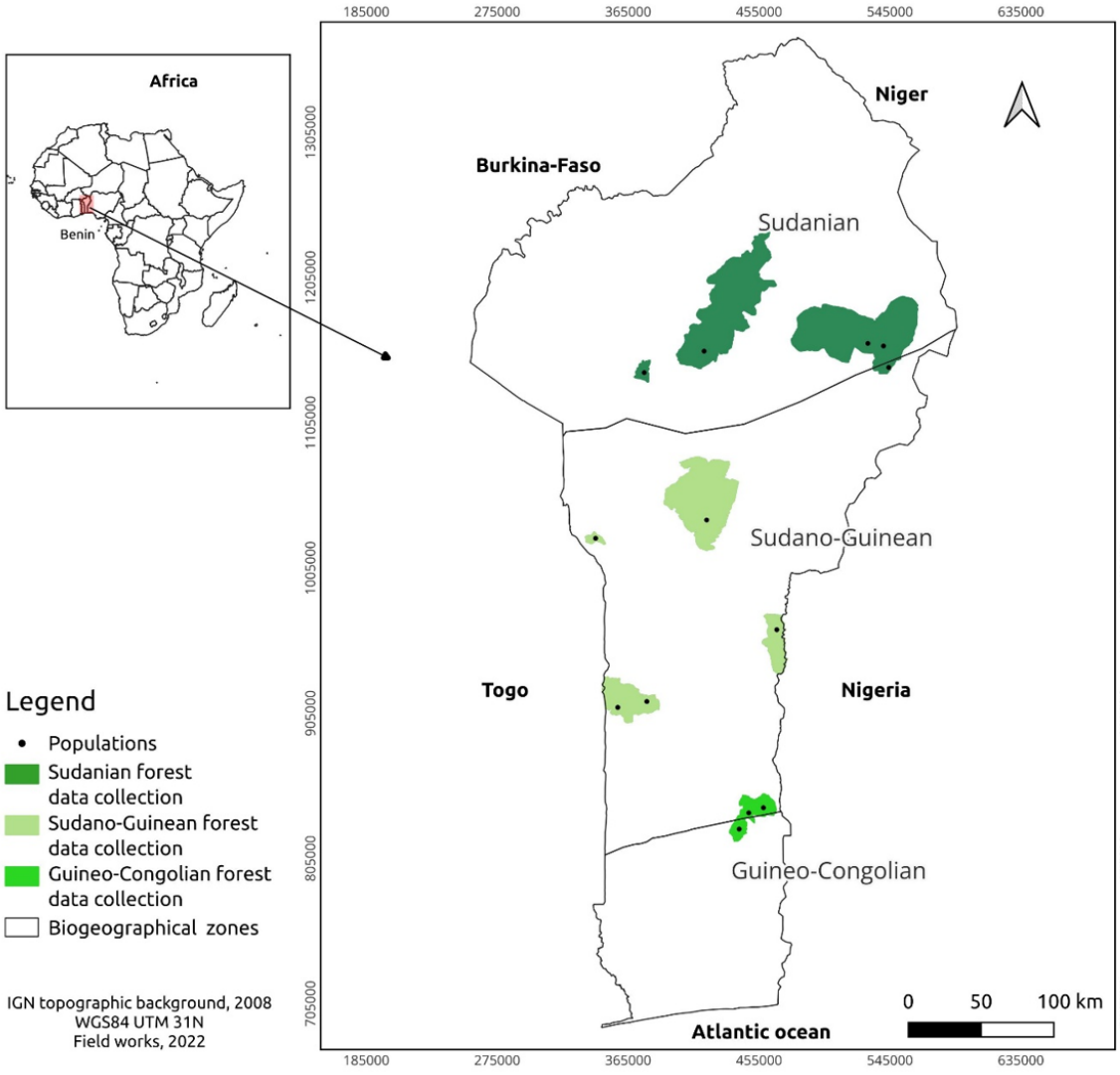


Figure 1. Study area showing the biogeographical presence of *Pseudocedrela kotschyi* in Benin and the data collection forests

reflects the southern distribution limit of the species, which is ecologically characteristic of the Sudanian region and does not extend further into the Guineo-Congolian domain (Table 1). Measurements included diameter at breast height (DBH) and total height, assessed with appropriate instruments. Sampling involved ten leaflets and up to 30 mature fruits per tree, along with three to five seeds per fruit (Figure 2). Leaflet dimensions were measured using a graduated ruler, while fruit and seed length, width and thickness were recorded using precise electronic callipers, with weights recorded using a precision balance (Figure 2). The total number of seeds per fruit and per individual seed was documented. The measured traits are presented in Table 2.

Data analysis

The morphological characteristics (fruit and leaf) of trees in populations from each biogeographical zone were assessed using descriptive statistics, including means, standard errors and coefficients of variation. Prior to inferential analyses, the assumptions of normality and homogeneity of variances were verified. Normality of residuals was tested using the Shapiro-Wilk test (Shapiro and Wilk, 1965), while homogeneity of variances was assessed using Levene's test. As these assumptions were satisfied, differences in morphological traits among biogeographical zones were evaluated using a one-way analysis of variance (ANOVA). When a significant effect was detected ($p < 0.05$), mean

Table 1. Number of samples per biogeographical zones

Biogeographical zones	No. of populations	No. of individuals sampled	No. of fruits	No. of leaves
Guineo-Congolian	3	59	1,412	1,312
Sudano-Guinean	5	54	1,083	1,193
Sudanian	5	37	591	346

Figure 2. Example a) seeds, b) fruits and c) leaves of *Pseudocedrela kotschyi*

Table 2. Mean values, standard error of the mean, and coefficient of variation of the morphological characteristics of *Pseudocedrela kotschyi* fruits and leaves across biogeographical zones in Benin. On the same line and for each character, values with the same letters are not statistically different (Student-Newman-Keuls test). Df, degrees of freedom; F, F value; Prob, probability; N, number of samples; se, standard error; CV, coefficient of variation; GC, Guineo-Congolian; S, Sudanian; SG, Sudano-Guinean.

Quantitative descriptors		GC	S	SG	Df	F	Prob
Fruits	N	1,412	591	1,083	2, 3083		
Large seed width (cm)	mean	1.02[b]	0.96[a]	1.01[b]		23.03	< 0.001
	se	0	0.01	0.01			
	CV	0.17	0.25	0.18			
Seed length (cm)	mean	4.19[c]	3.86[a]	4.01[b]		41.58	< 0.001
	se	0.02	0.03	0.02			
	CV	0.17	0.22	0.19			
Fruit length (cm)	mean	10.27[b]	9.5[a]	9.47[a]		46.25	< 0.001
	se	0.05	0.08	0.08			
	CV	0.19	0.21	0.28			
Number of seeds per fruit	mean	21.00[c]	18.00[a]	20.00[b]		64.39	< 0.001
	se	0.12	0.25	0.17			
	CV	0.21	0.33	0.28			
Seed weight (g)	mean	0.56[b]	0.46[a]	0.69[c]		103.6	< 0.001
	se	0.01	0.01	0.01			
	CV	0.58	0.62	0.5			
Fruit weight (g)	mean	51.59[c]	40.17[a]	46.66[b]		105.5	< 0.001
	se	0.39	0.73	0.53			
	CV	0.28	0.44	0.38			
Leaves	N	1312	346	1193	2, 2848		
Leaflet width (cm)	mean	4.13[b]	3.85[a]	4.14[b]		4.233	0.1193
	se	0.05	0.14	0.04			
	CV	0.40	0.66	0.36			
Leaf length (cm)	mean	36.24[c]	27.78[a]	33.57[b]		163.5	< 0.001
	se	0.22	0.38	0.22			
	CV	0.22	0.26	0.23			
Leaflet length (cm)	mean	10.56[b]	8.97[a]	10.32[b]		40.2	< 0.001
	se	0.08	0.14	0.09			
	CV	0.27	0.29	0.30			

comparisons were performed using the Student-Newman-Keuls (SNK) post-hoc test (Newman, 1939; Keuls, 1952). The SNK test was selected due to its suitability for detecting ordered differences among groups along ecological gradients and its effectiveness with large sample sizes, allowing a biologically meaningful interpretation of gradual morphological variation across biogeographical zones.

GPS coordinates of each sampled individual were used to extract values for 19 bioclimatic variables from the WorldClim database (http://www.worldclim.org, accessed on March 25, 2022). These bioclimatic data, representing the current climate (1970–2000), were obtained at a spatial resolution of 30 arc-seconds (Fick and Hijmans, 2017). Multicollinearity among the 19 variables (Merow *et al*, 2013) was tested based on a sample matrix of occurrence points and bioclimatic variables using the variance inflation factor (VIF) with the SDM package in R 3.6.3 (R Core Team, 2020) (Naimi and Araújo, 2016). A VIF threshold of < 10 was applied, following commonly used guidelines in ecological studies (Dormann *et al*, 2013), particularly for exploratory multivariate analyses such as principal component analysis (PCA). This threshold allowed the retention of ecologically relevant variables while limiting excessive collinearity, resulting in the selection of seven bioclimatic variables for further analysis. These are variables bio2 (mean diurnal range), bio3 (isothermality), bio5 (maximum temperature of the warmest month), bio10 (mean temperature of the warmest quarter), bio11 (mean temperature of the coldest quarter), bio12 (annual precipitation) and bio14 (precipitation of the driest month).

To identify correlations between fruit morphological traits and bioclimatic variables across biogeographical zones, PCA was conducted using the FactoMineR and factoextra packages, with prior standardization of the data. The PCA was applied to the mean values of various fruit morphological traits and corresponding bioclimatic variables

by latitude to emphasize large-scale ecological gradients and reduce local-scale variability. This approach was used for exploratory purposes only and not for statistical inference. Individual-level data were retained for all hypothesis-testing analyses to avoid pseudo-replication and to preserve within-population variability. This analysis was complemented by a correlogram, which assessed pairwise relationships between morphological and bioclimatic variables using Pearson's correlation coefficients. Statistical significance of correlations was evaluated based on associated p-values. To identify distinct fruit morphotypes, a hierarchical cluster analysis (HCA) was performed using the HCPC function of the FactoMineR package, which first applies a PCA to standardized fruit morphological traits and then conducts hierarchical clustering using Ward's minimum variance method based on Euclidean distances calculated from the retained principal components. The resulting clusters were further analyzed through canonical discriminant analysis (CDA) to evaluate the degree of differentiation among morphotypes, using the candisc function within the same package. For each identified morphotype, the mean, standard error, and coefficient of variation were calculated for all morphological traits to describe their distinguishing characteristics. To explore the association between morphotypes and biogeographical zones, a multiple correspondence analysis (MCA) was performed using the FactoMineR package. All statistical analyses were carried out in R version 3.6.3 (R Core Team, 2020).

Results

Morphological variability of *P. kotschyi* according to biogeographical zones

Fruit morphological traits varied significantly across zones ($p < 0.001$), as did the morphological traits of the leaves ($p < 0.001$), except for leaflet width ($p = 0.1193$) (Table 2). The highest values for several traits, such as seed number (21.00 ± 0.12), seed length (4.19 ± 0.02cm), fruit weight (51.59 ± 0.39g) and leaf length (36.24 ± 0.22cm), were recorded in the Guineo-Congolian zone, followed by the Sudano-Guinean zone (Table 2). The heaviest seeds (0.69 ± 0.01g) were found in the Sudanian-Guinean zone, followed by the Guineo-Congolian zone. Seed width (1.02 ± 0cm) and leaflet length (10.56 ± 0.08cm) had the highest values in the Guineo-Congolian and Sudano-Guinean zones. The longest fruits (10.27 ± 0.05cm) were also recorded in the Guineo-Congolian zone (Table 2). The coefficient of variation (CV) indicated significant variability among the different morphological traits of fruits and leaves, ranging from 17% for maximum seed width and seed length to 58% for seed weight (Table 2).

Influence of climatic factors on the quantitative morphological characteristics of the fruits of *P. kotschyi*

PCA was performed on standardized morphological traits and bioclimatic variables in order to explore their joint variation and identify the main gradients structuring the data. Morphological traits were strongly associated with Axis 1, while most bioclimatic variables – except for bio2 and bio14 – align with Axis 2. Axis interpretation was based on eigenvalues > 1, variance explained, and variable loadings, with absolute loading values ≥ 0.50 considered meaningful. Axis 1 was primarily defined by high positive loadings of fruit traits, heavier and longer fruits with more seeds, mainly influenced by latitude, bio2, bio3 and bio14. Axis 2 represents climatic variation, driven by bio5, bio10, bio11 and bio12. The Sudanian zone aligns with latitude and bio2, while the Sudano-Guinean zone is influenced by bio3, bio12 and bio14. The Guineo-Congolian zone, on the other hand, exhibits the best fruit characteristics, except for seed weight, which is associated with the variables maximum temperature of the warmest month (bio5), mean temperature of the warmest quarter (bio10), and mean temperature of the coldest quarter (bio11) (Figure 3). The correlogram revealed that only a limited number of correlations between morphological traits and bioclimatic variables were statistically significant, and most exhibited moderate effect sizes ($|r| < 0.5$). However, the correlogram of morphological traits and climatic variables did not provide strong or unambiguous statistical evidence of robust relationships between the two groups of variables (Figure 4).

Identification and discrimination of three distinct fruit morphotypes of *P. kotschyi*

Hierarchical classification identified three distinct fruit morphotypes (Figure 5). Figure 5 presents the projection of the three morphotypes in a coordinate system (first Wilks' Lambda function = 0.101, $p < 0.001$; second Wilks' Lambda function = 0.440, $p < 0.001$) represented by two canonical discriminant axes, which retain 100% of the initial information. Morphotypes 1 and 2 are clearly separated based on the measured morphological traits, as are morphotypes 1 and 3 (Figure 6).

Morphotype 1 is characterized by the lowest values for fruit morphological traits (Table 3). Morphotype 2 includes long fruits (10.81 ± 0.22cm) with high variability and the highest number of seeds (23 ± 0.14) with low variation. Morphotype 3 comprises heavy (0.99 ± 0.05g) and long seeds (4.46 ± 0.12cm) with little variation. Morphotypes 2 and 3 have the heaviest fruits (Table 3).

Pairwise correlation between morphological traits and biogeographical zones

The relationships between fruit morphotypes and biogeographical zones were statistically significant (chi-square = 279.87, df = 4, p-value < 2.2e-16). The MCA captured 60.45% of the information on the first two axes (Figure 7). The Sudanian and Sudano-Guinean zones, along with morphotype 3, were the most represented variables on axis 1, while the Guineo-Congolian zone and morphotypes 1 and 2 were well represented on axis 2. Morphotype 1 was most frequent in the Sudanian zone, whereas morphotype 2 was primarily found in the Guineo-Congolian zone, and morphotype 3 is dominant in the Sudano-Guinean zone.

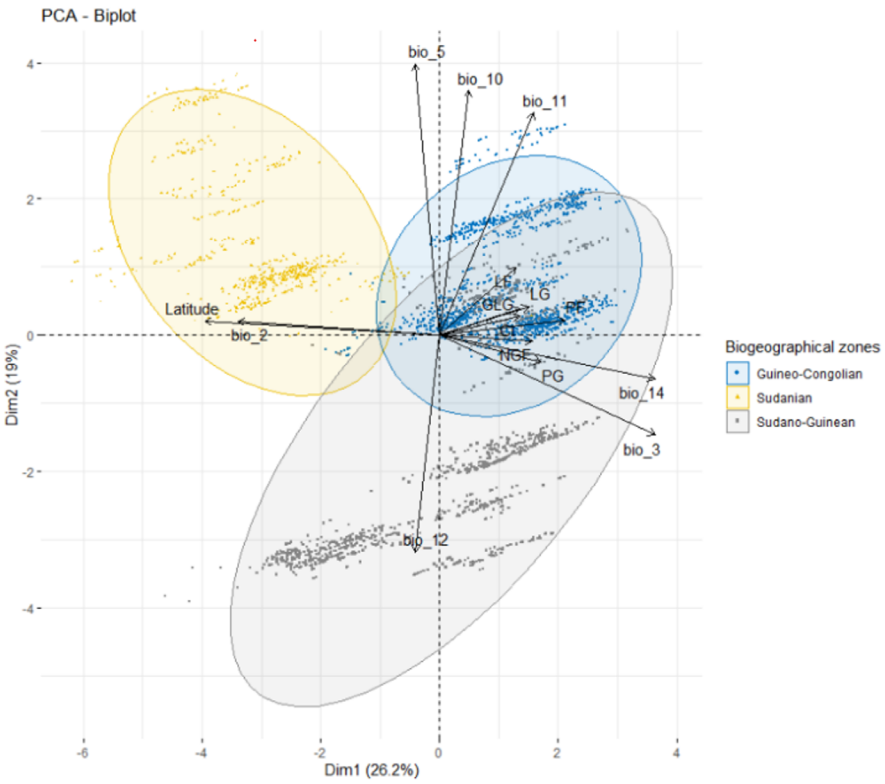


Figure 3. PCA biplot showing the distribution of *Pseudocedrela kotschyi* individual trait values within the axis system defined by seven bioclimatic variables and latitude. The ellipses represent 95% confidence envelopes for climatic zones. LF, fruit length (cm); PF, fruit weight (g); NGF, number of seeds per fruit; LG, seed length (cm); GLG, Seed width (cm); PG, seed weight (g); bio2, mean diurnal range; bio3, isothermality; bio5, maximum temperature of the warmest month; bio10, mean temperature of the warmest quarter; bio11, mean temperature of the coldest quarter; bio12, annual precipitation; bio14, precipitation of the driest month.

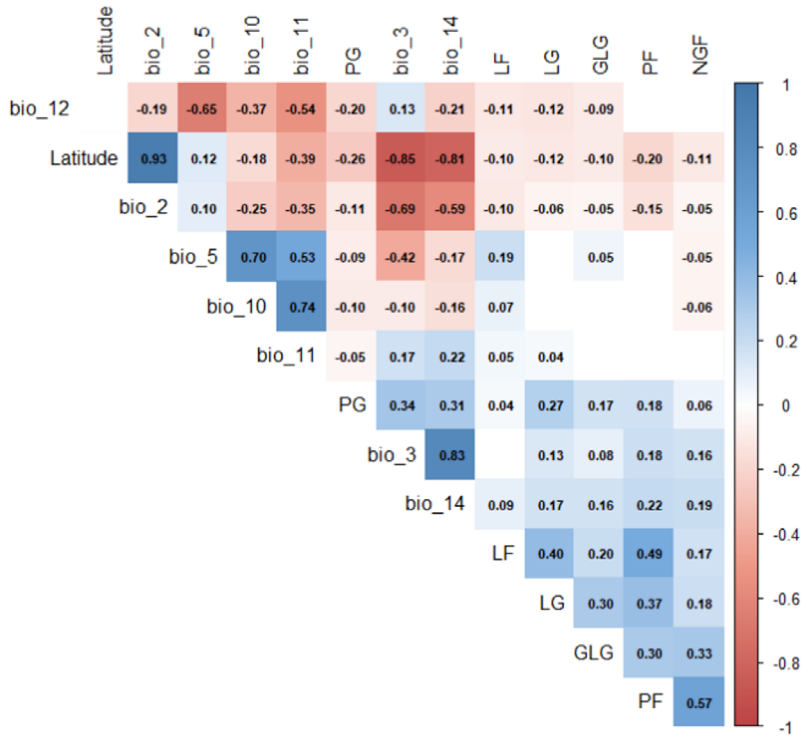


Figure 4. Correlogram of fruit morphological and climatic variables. The colour palette indicates positive (blue) and negative (red) 2 correlation coefficients. A positive coefficient indicates that the two variables move in the same direction, and a negative coefficient indicates the opposite. LF, fruit length (cm); PF, fruit weight (g); NGF, number of seeds per fruit; LG, seed length (cm); GLG, seed width (cm); PG, seed weight (g); bio2, mean diurnal range; bio3, isothermality; bio5, maximum temperature of the warmest month; bio10, mean temperature of the warmest quarter; bio11, mean temperature of the coldest quarter; bio12, annual precipitation; bio14, precipitation of the driest month.

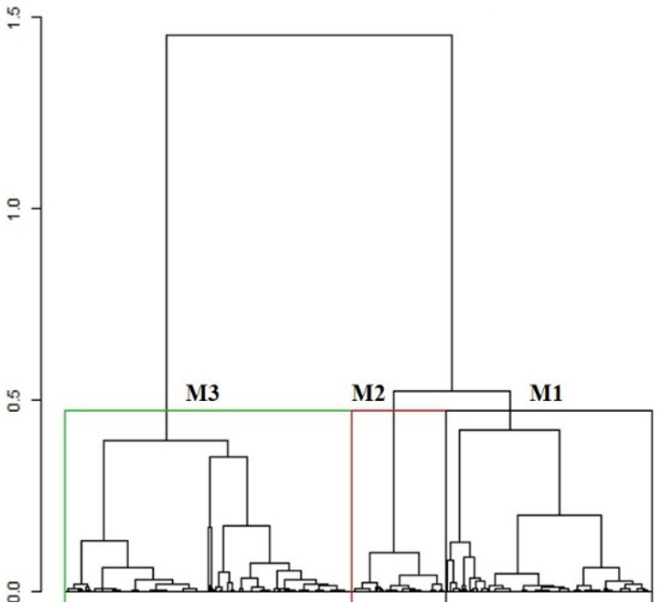


Figure 5. Dendrogram showing the different fruit morphotypes of *Pseudocedrela kotschyi*, based on individual values of fruit traits sampled in Benin. M1, morphotype 1; M2, morphotype 2; M3, morphotype 3.

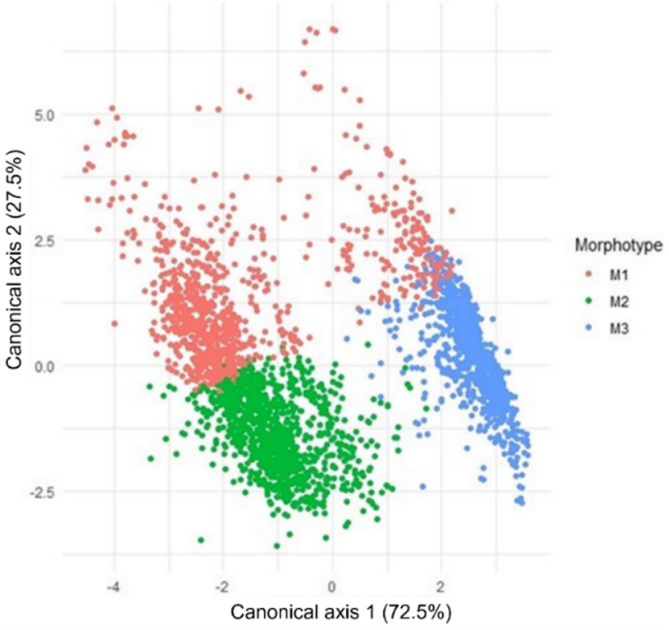


Figure 6. Projection of fruit morphotypes of *Pseudocedrela kotschyi*, based on individual values of fruit traits sampled in Benin formed by the first two canonical axes. M1, morphotype 1; M2, morphotype 2; M3, morphotype 3.

Table 3. Characteristics of three identified fruit morphotypes observed in *Pseudocedrela kotschyi* populations sampled in Benin. In the same row and for each trait, values sharing the same letters are not statistically different (Student-Newman-Keuls test). Prob, probability; mean, average; cv, coefficient of variation; M1, morphotype 1; M2, morphotype 2; M3, morphotype 3.

Quantitative descriptors		M1	M2	M3	Prob
Large seed width (cm)	mean	0.86[a]	1.07[b]	1.06[b]	< 0.001
	CV	0.24	0.13	0.14	
Seed length (cm)	mean	3.53[a]	4.13[b]	4.46[c]	< 0.001
	CV	0.25	0.14	0.12	
Fruit length (cm)	mean	8.28[a]	10.81[c]	10.18[b]	< 0.001
	CV	0.21	0.22	0.18	
Number of seeds per fruit	mean	15[a]	23[c]	21[b]	< 0.001
	CV	0.38	0.14	0.16	
Seed weight (g)	mean	0.43[b]	0.34[a]	0.99[c]	< 0.001
	CV	0.66	0.31	0.05	
Fruit weight (g)	mean	31.05[a]	54.72[b]	54.82[b]	< 0.001
	CV	0.39	0.22	0.26	

Discussion

Morphological variation of *P. kotschyi* across biogeographical zones of Benin

This study assessed the variability of leaf, fruit, and seed morphological traits of *P. kotschyi* and examined the role of environmental variables in the observed morphological patterns across the three biogeographical zones of Benin. The morphological characteristics of *P. kotschyi* fruits and leaves varied significantly between biogeographical zones. This variability reveals heterogeneity of *P. kotschyi* populations across distinct biogeographical regions.

The results indicated that all traits exhibited significant variation (CV > 17%). This pronounced variability of morphological traits around their mean values highlights the diversity within *P. kotschyi* populations. Similar patterns of intraspecific variation have been reported in other multipurpose woody species, such as *Pterocarpus erinaceus* (Konda *et al*, 2025) and *Strychnos spinosa* (Avakoudjo *et al*, 2021), which also show considerable morphological diversity across their distribution ranges. Such variation is often

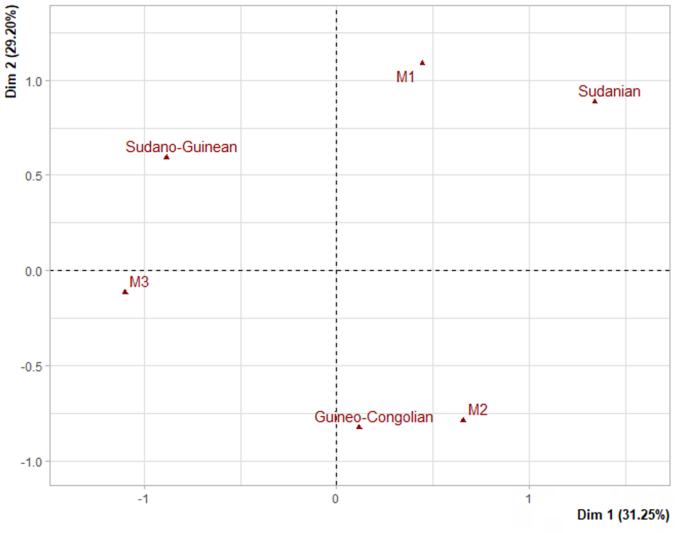


Figure 7. Graphical representation of the results from the multiple correspondence analysis (MCA), showing the distribution of biogeographical zones and fruit morphotypes. M1, morphotype 1; M2, morphotype 2; M3, morphotype 3.

attributed to the heterogeneous environmental conditions under which these species grow.

The highest values for morphological traits were generally observed in the Guineo-Congolian zone. Traits such as seed number, seed length, fruit length, fruit weight and leaf length had the highest values in the Guineo-Congolian zone, followed by the Sudano-Guinean zone. Additionally, maximum seed width and leaflet length were recorded in both the Guineo-Congolian and Sudanian-Guinean zones. Other studies have also reported greater morphological trait values in the Guineo-Congolian zone, including those by Daï *et al* (2024) and Houehanou *et al* (2019). This trend may be attributed to the higher water availability in this area, where annual rainfall ranges from 950 to 1,400mm (Adomou *et al*, 2006).

The high seed weight observed in the Sudano-Guinean zone could be explained by the presence of an ecological optimum. This zone likely offers a favourable combination of factors, such as temperature, seasonal humidity and soil structure; this promotes the development of heavier seeds, which are advantageous for seedling survival. For example, a study on *Prosopis africana* in Benin demonstrated that fruit traits (including seed weight) were significantly adapted to climatic zones, with the highest values recorded in areas with the most favourable ecological conditions (coefficient of variation up to 58%) (Towanou *et al*, 2015). Similarly, the Sudano-Guinean zone may represent an ecological optimum for *P. kotschyi*, where environmental conditions maximize seed weight, even though this trait does not follow the same pattern as other morphological traits.

Furthermore, this observation may be attributed to the presence of the species in agroecosystems of the Guineo-Congolian zone, where it is relatively more protected due to its utility compared to the other two zones (Deguenonvo *et al*, 2024). Based on these findings, we conclude that the Guineo-Congolian zone would be the most suitable for a conservation and domestication programme of *P. kotschyi* in Benin.

Influence of climatic and environmental factors on morphological traits of *P. kotschyi*

The study found no strong correlation between morphological traits and bioclimatic variables. This may be partly explained by the species' distribution range, which does not fully extend across all latitudes of the Guineo-Congolian zone. Indeed, *P. kotschyi* is ecologically classified as a Sudanian species, and individuals sampled in the so-called Guineo-Congolian zone were located near the transitional boundary with the Sudano-Guinean zone. Therefore, the limited latitudinal range of occurrence in the southern zone may restrict the expression of morphological responses to broader climatic gradients. Similar observations have been made in other studies, where the ecological amplitude of species influences their trait-environment relationships (Assogbadjo *et al*, 2011). Moreover, integrating additional variables – such as genetic diversity and soil properties - could provide a more comprehensive understanding of the factors driving morphological variability (Freschet *et al*, 2017).

The low values observed for morphological traits, such as small seeds and small fruits in Sudanian zone, could be attributed to hot climates, which may affect their overall quality and viability. At higher latitudes, fruit and seed characteristics tend to exhibit lower values. High latitudes often result in reduced fruit and seed sizes, as observed in various species, including *Cynodon dactylon*, where latitude significantly influences morphological traits (Zhang *et al*, 2018). Moreover, in the Guineo-Congolian zone, where temperature conditions are optimal, *P. kotschyi* tends to produce larger and heavier fruits. This finding aligns with the results of Hounkpèvi *et al* (2016) on *Vitex doniana* Sweet.

Relationship between fruit morphotypes and biogeographical zones

The identification and classification of fruit morphotypes are essential steps in understanding plant diversity and adaptation across different biogeographical zones. The analysis of *P. kotschyi* fruit morphological traits allowed the distinction of three morphotypes, each associated with specific climatic conditions. Hierarchical classification revealed three distinct types based on morphological traits. These morphotypes differed in characteristics such as fruit length, seed number and seed weight, which are essential for plant reproduction and survival. Discriminate analysis confirmed the distinctive nature of these morphotypes, highlighting the role of morphological traits in ecological and evolutionary processes. This classification aligned with findings from other studies on fruit and seed morphometry, which highlighted the importance of these traits in plant adaptation and evolutionary processes (Adjacou *et al*, 2024; Houndonougbo *et al*, 2020).

Morphotype 1 was characterized by the lowest values of fruit morphological traits, indicating potential adaptation to more arid conditions with a high diurnal temperature range (bio2) and the influence of latitude. This aligns with the observed trend where higher latitudes lead to reduced fruit and seed sizes (Zhang *et al*, 2018). Morphotype 2 consisted of longer fruits with significant variation in length, possibly reflecting a strategy that enhances seed dispersal and optimal germination (Avakoudjo *et al*, 2021; Lawin *et al*, 2021; Daï *et*

al, 2024). Morphotype 3 included heavy and long seeds, and high biomass production (Houndonougbo *et al*, 2020), which benefit seedling establishment and are likely to promote good germination and growth of *P. kotschyi* seedlings, as reported for *Uvaria chamae* by Daï *et al* (2024) and for *Cola millenii* by Lawin *et al* (2021). Heavier seeds may improve seedling survival by enhancing resilience to rainfall variability and specific edaphic conditions (Freschet *et al*, 2017).

Perspectives for the domestication and sustainable management of *P. kotschyi* in Benin

The results indicate that the Guineo-Congolian zone presents favourable conditions for the conservation and potential domestication of *P. kotschyi*, particularly for objectives related to vegetative growth and potential timber production. This zone provides higher water availability, moderate temperatures and enhanced protection of individuals within agroecosystems, which may support better vegetative growth and timber yield (Deguenonvo *et al*, 2024). However, when considering domestication for fruit or seed production, the Sudano-Guinean zone may offer more optimal conditions. This is supported by the observation that seed weight - an important trait for reproductive success and seedling establishment – was significantly higher in this zone, possibly reflecting an ecological optimum for reproductive performance.

Although *P. kotschyi* seeds are not currently utilized in human food systems, their potential for future industrial applications, such as medicinal use or oil extraction, may increase interest in their production. Consequently, both biogeographical zones offer strategic value depending on domestication objectives: the Guineo-Congolian zone is better suited for timber production and conservation, while the Sudano-Guinean zone presents greater potential for enhancing fruit traits and seed valorization. To support *in situ* and *ex situ* conservation strategies, we recommend the following actions: (1) preserve natural populations in the Guineo-Congolian zone, (2) investigate genetic diversity to better understand adaptation mechanisms and domestication potential, and (3) engage local communities actively in conservation efforts by integrating them into management programmes to enhance sustainability.

Conclusion

This study reveals notable morphological variation in *P. kotschyi* across Benin's biogeographical zones, influenced by environmental factors such as rainfall, temperature and latitude. Three distinct fruit morphotypes were identified, illustrating the species' adaptability to diverse climates. The Guineo-Congolian zone proved most favourable, producing larger and heavier fruits and seeds that enhance dispersal and seedling survival. However, the weak correlation between morphological traits and climatic variables suggests that other factors, particularly soil conditions and genetic diversity, may also shape this variability. These insights underscore the importance of adopting an integrated approach to conservation and domestication. Combining morphological, ecological and genetic data is essential to understanding the species' adaptive potential. Sustainable strategies, including protecting natural populations, selecting productive morphotypes and involving local communities, are crucial for ensuring the long-term survival of *P. kotschyi*. Such approaches will contribute to biodiversity conservation and the development of climate-resilient agroforestry systems in Benin.

Acknowledgements

The authors would like to thank the Alexander von Humboldt Foundation for its support for this study through equipment grant 3.4-8151/Houehanou (GA-Nr.).

Author contributions

TAGD and TDH Houehanou conceived the research. TAGD, DMA and TDH developed the methodology. TAGD, TDH, RS and FEDS conducted formal analysis and investigation. TAGD, DMA, RI and TDH wrote the original draft. TDH acquired the funding, provided the resources, and supervised the research. All authors participated in the review and editing of the manuscript.

Data availability statement

Accession-level data are available from the corresponding author upon reasonable request.

Conflict of interest statement

The authors have no relevant financial or non-financial interests to disclose.

References

Adjacou, D.M., Houehanou, D.T., Gouwakinnou, N.G., Prinz, K., Moussa, T., Mama, A.-R., Hellwig, F., Natta, K.A., 2024. Local knowledge and morphological variations in local landraces Mangifera indica L. in Northern Benin (West Africa). *Afr. J. Agric. Res.* 20, 650–666. https://doi.org/10.5897/AJAR2024.16678

Adjacou, D.M., Houehanou, T.D., Gouwakinnou, G.N., Natta, A.K., 2022. Connaissances ethnoécologiques des variétés locales de *Mangifera indica* L. dans l'Atacora au Bénin : usages, diversité et perceptions du changement climatique. *Ann. UP, Série Sci. Nat. Agron.* 12, 15–28. https://doi.org/10.56109/aup-sna.v12i1.107

Adomou, A.C., Sinsin, B., van der Maesen, L.J.G., 2006. Notulae florae beninensis 12: phytosociological and chorological approaches to phytogeography: a meso-scale study in Benin. *Systematics and geography of plants* 155–178.

Akoègninou, A., Van der Burg, W.J., Van der Maesen, L.J.G., 2006. Flore analytique du Bénin. Backhuys Publishers.

Alhassan, A.M., Ahmed, Q.U., Malami, I., Zakaria, Z.A., 2021. Pseudocedrela kotschyi: a review of ethnomedicinal uses, pharmacology and phytochemistry. Pharmaceutical Biology 59, 953–961. https://doi.org/10.1080/13880209.2021.1950776

Arbonnier, M., 2019. Arbres, arbustes et lianes d'Afrique de l'Ouest: Nouvelle édition 2019. Quae.

Assédé, E.P., Adomou, A.C., Sinsin, B., 2012. Relationship between stand regime and population structure of *Pseudocedrela kotschyi* (Meliaceae) and *Terminalia macroptera* (Combretaceae) in the Biosphere Reserve of

Pendjari (Benin, West Africa). I*nternational Journal of Biodiversity and Conservation* 4, 427–438. doi: https://doi.org/10.5897/IJBC12.028

Assogbadjo, A.E., Glegrave, R., Azihou, A.F., Kyndt, T., Codjia, J.T.C., 2011. Ethnic differences in use value and use patterns of the threatened multipurpose scrambling shrub (*Caesalpinia bonduc* L.) in Benin. *Journal of Medicinal Plants Research* 5, 1549–1557. http://hdl.handle.net/1854/LU-1256575

Avakoudjo, H.G.G., Idohou, R., Salako, K.V., Hounkpèvi, A., Koné, M.W., Assogbadjo, A.E., 2021. Diversity in tree and fruit traits of *Strychnos spinosa* Lam. along a climatic gradient in Benin: a step towards domestication. *Genet Resour Crop Evol* 68, 2423–2440. https://doi.org/10.1007/s10722-021-01140-5

Daï, E.H., Salako, K.V., Hotes, S., Assogbadjo, A.E., 2024. Morphological variability of 'bush banana' (*Uvaria chamae*) and its environmental determinants in Benin, West Africa. *Genet Resour Crop Evol* 71, 4049–4065. https://doi.org/10.1007/s10722-024-01926-3

Deguenonvo, A.G., Dossou, J., Idohou, R., 2020. Aptitude à la multiplication de *Pseudocedrela kotschyi* (Schweinf.) Harms par graines et par boutures de tige et de racine au Bénin. *International Journal of Biological and Chemical Sciences* 14, 2506–2516. https://doi.org/10.4314/ijbcs.v14i7.11

Deguenonvo, T.A.G., Adjacou, D.M., Idohou, R., Sodedja, R., Sobakin, F.E.D., Houehanou, T.D., Gouwakinnou, G.N., Natta, A.K., 2024. Synergizing climate dynamics, species distribution, and structural parameters for sustainable management of *Pseudrocedrela kotschyi* in Benin (West Africa). *Global Ecology and Conservation* 56, e03322. https://doi.org/10.1016/j.gecco.2024.e03322

Deguenonvo, T.A.G., Houehanou, T.D., Adjacou, D.M., Sobakin, F.E.D., Sodedja, R., Gouwakinnou, G.N., Natta, A.K., 2023a. Effet de la zone climatique et de la perturbation humaine sur la composition floristique et la diversité des habitats de *Pseudocedrela kotschyi* au Benin : Effect of climatic zone and human disturbance on floristic composition and diversity of *Pseudocedrela kotschyi* habitats in Benin. *International Journal of Biological and Chemical Sciences* 17, 2299–2311. https://dx.doi.org/10.4314/ijbcs.v17i6.13

Deguenonvo, T.A.G., Houehanou, T.D., Idohou, R., Yehouenou, N., Gouwakinnou, G.N., Natta, A.K., 2023b. Uses, Cultural Importance, and Fire Threat to *Pseudocedrela kotschyi* (Meliaceae): Evidence for the Availability Hypothesis in Benin (West Africa). *Economic Botany* 1–19. doi: https://doi.org/10.1007/s12231-023-09581-y

Diarra, M.L., Mariko, M., Mbaye, M.S., Noba, K., 2016. Plantes médicinales utilisées dans le traitement traditionnel du paludisme à Bamako (Mali). *International Journal of Biological and Chemical Sciences* 10, 1534–1541. doi: http://dx.doi.org/10.4314/ijbcs.v10i4.7

Dormann, C.F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., Marquéz, J.R.G., Gruber, B., Lafourcade, B., Leitão, P.J., Münkemüller, T., McClean, C., Osborne, P.E., Reineking, B., Schröder, B., Skidmore, A.K., Zurell, D., Lautenbach, S., 2013. Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography* 36, 27–46. https://doi.org/10.1111/j.1600-0587.2012.07348.x

Dossa, L.O.S.N., Dassou, G.H., Adomou, A.C., Ahononga, F.C., Biaou, S., 2021. Dynamique spatio-temporelle et vulnérabilité des unités d'occupation du sol de la Forêt Classée de Pénéssoulou de 1995 à 2015 (Bénin, Afrique de l'Ouest). *Sciences de la vie, de la terre et agronomie* 9.

Fick, S.E., Hijmans, R.J., 2017. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *International journal of climatology* 37, 4302–4315. doi: https://doi.org/10.1002/joc.5086

Freschet, G.T., Valverde-Barrantes, O.J., Tucker, C.M., Craine, J.M., McCormack, M.L., Violle, C., Fort, F., Blackwood, C.B., Urban-Mead, K.R., Iversen, C.M., Bonis, A., Comas, L.H., Cornelissen, J.H.C., Dong, M., Guo, D., Hobbie, S.E., Holdaway, R.J., Kembel, S.W., Makita, N., Onipchenko, V.G., Picon-Cochard, C., Reich, P.B., de la Riva, E.G., Smith, S.W., Soudzilovskaia, N.A., Tjoelker, M.G., Wardle, D.A., Roumet, C., 2017. Climate, soil and plant functional types as drivers of global fine-root trait variation. *Journal of Ecology* 105, 1182–1196. https://doi.org/10.1111/1365-2745.12769

Grubben, G.J.H., 2008. Plant resources of tropical Africa (PROTA). Prota.

Houehanou, T.D., Prinz, K., Hellwig, F., Assogbadjo, A.E., Gebauer, J., Kakaï, R.L.G., Sinsin, B., 2019. Morphological trait variation and relationships of *Afzelia africana* Sm. caused by climatic conditions and anthropogenic disturbance in Benin (West Africa). G*enetic Resources and Crop Evolution* 66, 1091–1105. https://doi.org/10.1007/s10722-019-00773-x

Houehanou, T.D., Prinz, K., Koua, D., Hellwig, F., Ebou, A., Gouwakinnou, G., Assogbadjo, A.E., Glele Kakaï, R.L., Zézé, A., 2023. Genetic diversity and population structure of a threatened tree species *Afzelia africana* Sm. ex Pers. among climatic zones for conservation challenges in Benin (West Africa). *Genetic Resources and Crop Evolution* 1–16. doi: https://doi.org/10.1007/s10722-022-01523-2

Houndonougbo, J.S.H., Kassa, B., Salako, V.K., Idohou, R., Assogbadjo, A.E., Glèlè Kakaï, R., 2020. Perceived variation of fruit traits, and preferences in African locust bean [*Parkia biglobosa* (Jacq.) Benth.] in Benin: implications for domestication. *Genet Resour Crop Evol* 67, 1315–1329. https://doi.org/10.1007/s10722-020-00915-6

Hounkpèvi, A., Azihou, A.F., Kouassi, É.K., Porembski, S., Glèlè Kakaï, R., 2016. Climate-induced morphological variation of black plum (*Vitex doniana* Sw.) in Benin, West Africa. *Genet Resour Crop Evol* 63, 1073–1084. https://doi.org/10.1007/s10722-016-0409-9

Hounkpèvi, A., Salako, V.K., Donhouédé, J.C.F., Daï, E.H., Tovissodé, F., Glèlè Kakaï, R., Assogbadjo, A.E., 2020. Natural intraspecific trait variation patterns of the wild soursop *Annona senegalensis* (Annonaceae) along a climatic gradient in Benin, West Africa. *Plant Ecology and Evolution* 153, 455–465. doi: https://doi.org/10.5091/plecevo.2020.1576

Ikabanga, D.U., Stévart, T., Koffi, K.G., Monthé, F.K., Doubindou, E.C.N., Dauby, G., Souza, A., M'batchi, B., Hardy, O.J., 2017. Combining morphology and population genetic analysis uncover species delimitation in the widespread African tree genus *Santiria* (Burseraceae). *Phytotaxa* 321, 166–180. https://doi.org/10.11646/phytotaxa.321.2.2

Keuls, M., 1952. The use of the „studentized range" in connection with an analysis of variance. *Euphytica* 1, 112–122.

Konda, B., Dimobe, K., Salako, K.V., Dembélé, J.B., Boussim, I.J., 2025. Morphological variability of *Pterocarpus erinaceus* Poir. along a climate gradient in Burkina Faso, West Africa: implications for conservation and domestication. *Genet Resour Crop Evol.* https://doi.org/10.1007/s10722-024-02310-x

Lawin, I.F., Fandohan, A.B., Salako, K.V., Assogbadjo, A.E., Ouinsavi, C.A.I.N., 2021. Morphological variability of fruits of *Cola millenii* K. Schum. from seven phytogeographical

districts in Benin: opportunity for domestication. *Genetic Resources and Crop Evolution* 68, 1225–1242. doi: https://doi.org/10.1007/s10722-020-01086-0

Mars, M., Marrakchi, M., 2000. Étude de la variabilité intra-arbre chez le grenadier (*Punica granatum* L.): application à l'échantillonnage des fruits. *Fruits* 55, 347–355.

Merow, C., Smith, M.J., Silander Jr, J.A., 2013. A practical guide to MaxEnt for modeling species' distributions: what it does, and why inputs and settings matter. *Ecography* 36, 1058–1069. doi: https://doi.org/10.1111/j.1600-0587.2013.07872.x

Moussilimi, I.T., Koura, K., Aoudji, A.K.N., Gbetoho, J.A., Akouehou, G.S., Ganglo, J.C., 2022. Caractéristiques structurales et écologiques des populations de *Pseudocedrela kotschyi* de la forêt de Pénéssoulou (Bénin). *Annales de l'Université de Parakou - Série Sciences Naturelles et Agronomie* 12, 13–26. https://doi.org/10.56109/aup-sna.v12i2.122

Naimi, B., Araújo, M.B., 2016. sdm: a reproducible and extensible R platform for species distribution modelling. *Ecography* 39, 368–375. https://doi.org/10.1111/ecog.01881

Newman, D., 1939. The distribution of range in samples from a normal population, expressed in terms of an independent estimate of standard deviation. *Biometrika* 31, 20–30.

Ouattara, B., Sanou, L., Koala, J., Hien, M., 2022. Perceptions locales de la dégradation des ressources naturelles du corridor forestier de la Boucle du Mouhoun au Burkina Faso. *Bois & Forets des Tropiques* 352, 43–60. doi: https://doi.org/10.19182/bft2022.352.a36935

R. Core Team, 2020. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. url: https://www.R-project.org/.

Shapiro, S.S., Wilk, M.B., 1965. An analysis of variance test for normality (complete samples). *Biometrika* 52, 591–611.

Stévart, T., Dauby, G., Lowry, P.P., Blach-Overgaard, A., Droissart, V., Harris, D.J., Mackinder, B.A., Schatz, G.E., Sonké, B., Sosef, M.S.M., Svenning, J.-C., Wieringa, J.J., Couvreur, T.L.P., 2019. A third of the tropical African flora is potentially threatened with extinction. *Sci. Adv.* 5, eaax9444. https://doi.org/10.1126/sciadv.aax9444

Towanou, H., Charlemagne, G.D.S.J., Christine, O., Nestor, S., 2015. Morphological Variability of *Prosopis africana* (Guill., Perrott. Et Rich.) Taub in Benin, West Africa. *American Journal of Plant Sciences* 6, 1069–1079. https://doi.org/10.4236/ajps.2015.67111

White, F., 1983. The vegetation of Africa.

Yaoitcha, A.S., Aboh, A.B., Zoffoun, A.G., Houinato, M., Mensah, G.A., Sinsin, B., Akpo, E.L., 2016. Potentiel de régénération des chantiers de production du charbon de bois au Centre-Bénin. *International Journal of Biological and Chemical Sciences* 10, 1702–1716. https://doi.org/10.4314/ijbcs.v10i4.21

Yusuf, S.N.A., Rahman, A.M.A., Zakaria, Z., Subbiah, V.K., Masnan, M.J., Wahab, Z., 2020. Morphological Variability Identification of Harumanis Mango (*Mangifera indica* L.) Harvested from Different Location and Tree Age. *Trop Life Sci Res* 31, 107–143. https://doi.org/10.21315/tlsr2020.31.2.6

Zhang, J., Wang, M., Guo, Z., Guan, Y., Guo, Y., Yan, X., 2018. Variations in morphological traits of bermudagrass and relationship with soil and climate along latitudinal gradients. *Hereditas* 155, 31. https://doi.org/10.1186/s41065-018-0068-2

Zida, D., Sanou, L., Diawara, S., Savadogo, P., Thiombiano, A., 2020. Herbaceous seeds dominates the soil seed bank after long-term prescribed fire, grazing and selective tree cutting in savanna-woodlands of West Africa. *Acta Oecologica* 108, 103607. https://doi.org/10.1016/j.actao.2020.103607

Genetic Resources

ORIGINAL ARTICLE

Genetic Resources (2026), 7 (13), 89–102
DOI: 10.46265/genresj.OBAM5787
https://www.genresj.org
ISSN: 2708-3764

Assessment of phenotypic and genetic variability in Nepalese cucumber (*Cucumis sativus* L.) accessions

Pradip Thapa[a,*], Sandip Bohara[b], Subbechha Giri[c], Naturally KC[c], Basanta Kumar Rimal[c], Bal Krishna Joshi[a]

[a] *National Agriculture Genetic Resources Centre, Khumaltar, Lalitpur, Nepal*
[b] *Far Western University, Tikapur, Nepal*
[c] *Agriculture and Forestry University, Chitwan, Nepal*

Abstract: Nepal harbours a rich diversity of cucumber landraces adapted to diverse agroecological conditions, but their phenotypic diversity remains largely unexplored. Eleven cucumber accessions collected from ten districts of Nepal were characterized and evaluated at the research field of the National Agriculture Genetic Resources Centre. Twelve quantitative and 16 qualitative agromorphological traits were measured following the International Union for the Protection of New Varieties of Plants (UPOV) guidelines. Diversity was assessed using descriptive statistics, Shannon-Weaver diversity index, principal component analysis, clustering, heritability analysis and phenotypic path analysis. Qualitative traits exhibited low to very high diversity (H′ = 0.44–0.99), while quantitative traits showed low to high diversity (H′ = 0.21–0.71). The first three principal components explained nearly 70% of the total variation, with leaf length, days to first female flowering, days to first male flowering, fruit breadth, length of peduncle, fruit length, and total yield as major contributors. UPGMA clustering grouped the accessions into four clusters with 80% similarity level. Phenotypic path analysis indicated that fruit size, fruit breadth, fruit length, leaf width, distance between the nodes and peduncle thickness were key determinants of yield. High heritability was observed for days to first male flowering, peduncle length, and stem diameter. Accessions CO14392 and CO13634, characterized by high yield potential (14.98t/ha) and superior fruit dimensions, were identified as the most promising for breeding programmes. These findings provide a foundation for utilizing Nepalese cucumber landraces, with multi-location and multi-year evaluations, complemented by molecular characterization to fully exploit their potential.

Keywords: Agromorphological traits, Cluster dendrogram, Cucumber diversity, Landraces, PCA

Citation: Thapa, P., Bohara, S., Giri, S., KC, N., Rimal, B. K. and Joshi, B. K. (2026) "Assessment of phenotypic and genetic variability in Nepalese cucumber (*Cucumis sativus* L.) accessions", *Genetic Resources*, 7(13), pp. 89–102. doi: 10.46265/genresj.OBAM5787.

Introduction

Cucumber (*Cucumis sativus* L. 2n = 2x = 14), a member of the most genetically diverse family Cucurbitaceae, is an annual trailing vegetable grown in both outdoor fields and protected conditions (Cebeci *et al*, 2020; Raza *et al*, 2020; Sharma *et al*, 2020). The primary centre of origin of cucumber is India, where its domestication occurred about 3,000 years ago (Whitaker and Davis, 1962; Renner *et al*, 2007; Sebastian *et al*, 2010). Cucumber farming spread from India to secondary diversity centers, including China (Sebastian *et al*, 2010), Western Asia, Northern Africa and Southern Europe (Bisht *et al*, 2004; Lv *et al*, 2012). Although the genus *Cucumis* includes more than 50 species, based on geographical origin and chromosome number, the genus is broadly represented by two major cultivated lineages into African (*C. melo* L., 2n = 2x = 24) and Asian (*C. sativus* L., 2n = 2x = 14) types (Kirkbride, 1993; Sebastian *et al*, 2010; Ahmed *et al*, 2022). Cucumber is classified into four types according to the domestication process: *C. sativus* var. *hardwickii* (wild cucumber), *C. sativus* var. *xishuangbannanesis* (semi-wild cucumber), *C. sativus* var. *sikkimensis* (Sikkim cucumber),

* Corresponding author: Pradip Thapa (pradip.thapa876@gmail.com)

Received: 10.08.2025 | Accepted: 21.01.2026 | Published online: 26.02.2026

and *C. sativus* var. *sativus* (cultivated cucumber) (Yang *et al*, 2012; Weng, 2021). *C. hystrix* is the sister species to *C. sativus* (Sebastian *et al*, 2010), whereas *C. sativus* var. *hardwickii* (Royle) Alef. is a progenitor and wild relative of cultivated cucumber (Staub *et al*, 1997; Bisht *et al*, 2004), and shows variation in qualitative and quantitative parameters (Naegele and Wehner, 2016).

Globally, it is cultivated in 3.5 million hectares with a production of 178.03 million tonnes in 2023 (FAOSTAT, 2024). The diverse agroclimatic conditions found in Nepal offer peculiar opportunities for cucumber cultivation. The crop has tremendous economic significance in Nepal, with 9,463 hectares of land under its cultivation, yielding 150,213 tonnes, with a productivity of 15.58 t/ha annually (MoALD, 2025). Cucumber is used in various forms, including salads, pickles, fresh snacks, sauces, juices, and as an ingredient in cooked meals (Khulakpam *et al*, 2015; Shah *et al*, 2016). Nutritionally, cucumber pulp (100g) contains 11.6g ash, 15.9g protein, 0.13g fat, 6.77g crude fibre, 52.7g carbohydrates, 0.6g protein, 13.9g calcium, 7.8mg iron, 15.1g sodium, 43.7g potassium, 2.49mg copper, 5.27mg zinc, 0.79mg manganese, 7.08g magnesium (Niyi *et al*, 2019). Despite its low-calorie content, it contains essential vitamins, minerals, phytochemicals and antioxidants (Chakraborty and Rayalu, 2021). A natural compound found in large amounts in cucumbers, called Cucurbitacin B (CuB), mainly functions as an apoptosis inducer in a variety of human cancer cells (Gao *et al*, 2014).

Genetic variation is a prerequisite for the improvement of any crop (Gaikwad *et al*, 2011; Chalbi *et al*, 2023). Diversity assessment assists plant breeders in recognizing and selecting the desirable traits, minimizing the challenges associated with parental selection, thereby facilitating the progression of generations (Chikh-Rouhou *et al*, 2023; Azam *et al*, 2024). A landrace is a dynamic population of cultivated crops with historical existence, unique identity, genetic diversity, and often lacking systematic crop enhancement (Villa *et al*, 2005). It is regionally adapted to tolerate both biological and environmental stresses, leading to stable yield, and is connected to traditional agriculture practices (Zeven, 1998). A diverse range of cucumber landraces, along with their wild relatives, exists in Nepal (Mainali and Jyakhwa, 2023), and it is critical to gather and conserve the existing landraces to prevent genetic erosion.

The extensive cultivation of exotic hybrid cucumber varieties is a serious threat to landraces, which harbour crucial genes tolerant to both abiotic and biotic stresses (Ahmed *et al*, 2022). The National Genebank of Nepal preserves 69 local landraces (Genebank, 2023) and the National Seed Board (NSB), Nepal, has released one open-pollinated variety (Kusle) and registered 24 cucumber varieties (SQCC, 2024). The released variety Kusle is characterized by a high male-to-female ratio, extended growth cycle, and higher disease susceptibility, resulting in lower yield (Gautam *et al*, 2021).

Only a limited number of studies have focused on the characterization and evaluation of cucumber landraces in Nepal (Shakya *et al*, 2006; Mainali and Jyakhwa, 2023). Therefore, the primary purpose of conducting this study was the characterization and evaluation of Nepalese cucumber landraces. Through the use of multivariate data analysis (principal component analysis (PCA), cluster dendrogram, phenotypic and genetic variability and heritability analysis), this study aims to provide valuable insights into the phenotypic diversity and agricultural potential of these landraces to be valorized in breeding programmes.

Materials and methods

Plant materials and experimental site

Ten cucumber accessions, collected from ten districts of Nepal (Table 1), along with one check variety, Bhaktapur Local, were characterized and evaluated at the research field of the National Agriculture Genetic Resources Center (NAGRC) in Khumaltar, Lalitpur, in 2024. The experimental site is located at an altitude of 1,368m a.s.l., with coordinates 27°40'N latitude and 85°20'E longitude. The check variety Bhaktapur Local is the best commercial variety registered in 2018 (MoALD, 2024). The experiment region receives 4.4 to 105.7mm of rainfall and experiences average temperatures of 15.2ºC (min) and 28.2ºC (max) (Genebank, 2023)(Figure 1). The soil type at the research site was black loamy (Ghimire and Magar, 2017).

Table 1. Details of the accessions used in the experiment

Accession	Type	Collection district	Collection location		
			Altitude (m)	Latitude	Longitude
CO10293	Landrace	Gulmi	997	28.02	83.18
CO13538	Landrace	Panchthar	1,606	27.06	87.45
CO13634	Landrace	Baitadi	1,526	29.21	80.42
CO14242	Landrace	Jajarkot	1,706	28.41	82.13
CO14365	Landrace	Gorkha	1,101	27.59	84.39
CO14392	Landrace	Gorkha	1,063	27.58	84.40
CO14439	Landrace	Jumla	1,349	28.57	81.36
CO14589	Landrace	Sindhupalchowk	1,682	27.35	85.55
CO14765	Landrace	Nuwakot	1,005	27.54	85.05
SIRG-024 10CU	Landrace	Rasuwa	1,385	28.00	85.12
Bhaktapur Local	Commercial	Lalitpur	1,341	27.38	85.19

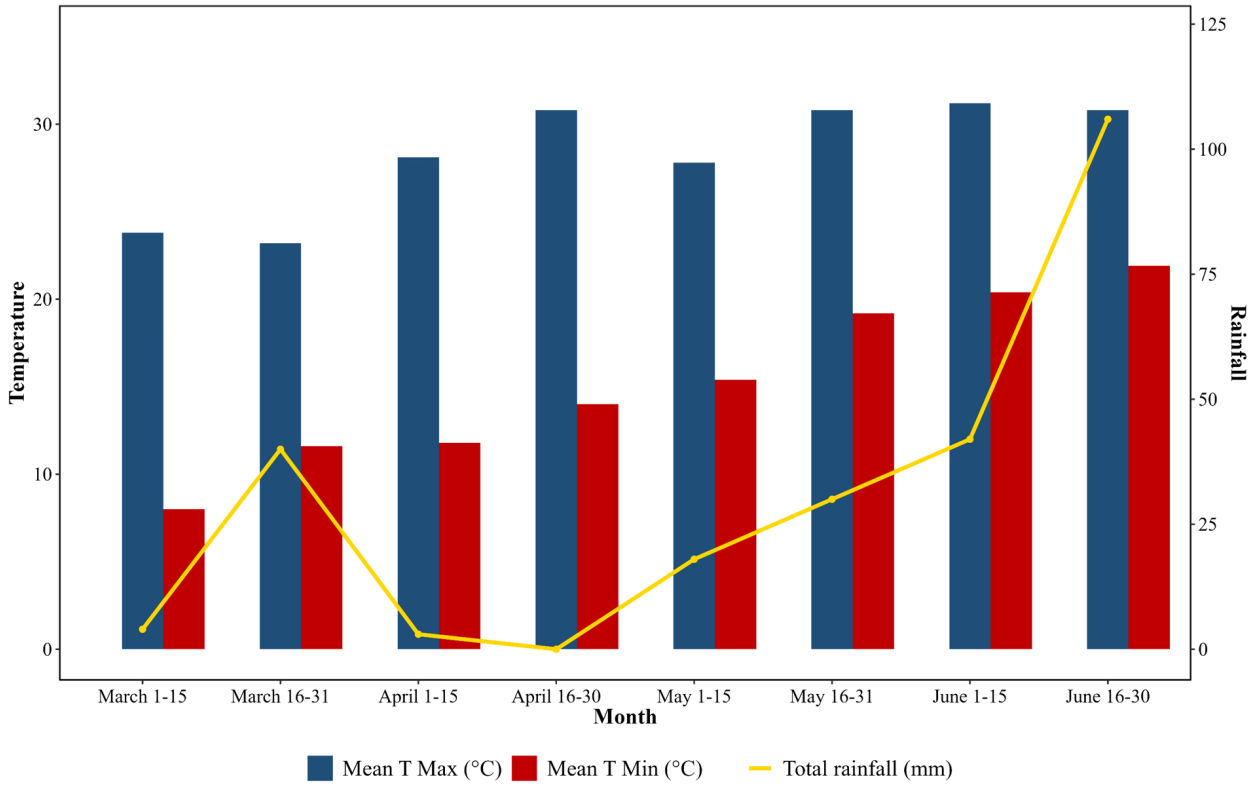


Figure 1. Average maximum and minimum temperature and total rainfall of the study site (Source: Genebank, 2024)

Seedling raising and transplanting

Seedlings were raised in polybags filled with farmyard manure (FYM) and soil mixture. Ten polybags were prepared, with three seeds sown per polybag. A total of 20 healthy seedlings were ultimately transplanted per accession per replication. The polybags were lightly irrigated and maintained under a low tunnel covered with a white polythene sheet for one month to ensure optimal temperature and rapid germination. After the emergence of true leaves (3–4 leaf stage), one-month-old seedlings were transplanted to the open field during late March to early April.

Experimental design

The accessions were planted in a randomized complete block design with three replications for agromorphological characterization and evaluation. An experiment was conducted from March to June 2024, using fertilizer at the rate of 140:40:100 NPK kg/ha along with 1.5t/ha FYM and recommended management practices (MoALD, 2024). The experimental plot measured 3m² (3m × 1m) with a spacing of 1.0m × 0.6m (RR × PP). Two seedlings were transplanted per hill. From each block, ten plants per accession were randomly selected for the measurement of quantitative traits.

Data collection

Morphological data of 16 qualitative (Supplemental Table 1) and 12 quantitative traits (Supplemental Table 2) were recorded using the standard descriptors for cucumber (UPOV, 2007). The qualitative traits were recorded based on the observation of descriptor states, whereas the quantitative traits were measured from selected sample plants.

Data analysis

The qualitative and quantitative data were recorded and entered into Microsoft Excel (2016). The quantitative data (Supplemental Table 3) were analyzed for descriptive statistics by using the pastecs package (Grosjean and Ibanez, 2002) in R Studio (4.4.3) (R Core Team, 2025). The qualitative data were analyzed by using the package summary tools (Comtois, 2014). Shannon's Weaver diversity index (H') (Shannon, 1948) was calculated by using the formula in Microsoft Excel. For both quantitative and qualitative traits, the data were classified into nine intervals, and the resulting diversity indices were categorized into four classes: low (0.1–0.4), intermediate (0.4–0.6), high (0.6–0.8), and very high (>0.8), as described by Eticha *et al* (2006). PCA was prepared by using the package factoextra (Kassambara and Mundt, 2016), factominor (Husson *et al*, 2006) and gridextra (Auguie, 2010). The estimates of genetic parameters – genotypic variance (GV), phenotypic variance (PV), genotypic coefficient of variance (GCV), phenotypic coefficient of variance (PCV), heritability (H^2_{bs}), genetic advance as a percentage of mean (GAM) – were calculated by using the variability package (Popat *et al*, 2020), whereas phenotypic path analysis was done using metan package (Olivoto and Lúcio, 2020) in R Studio. The optimum number of clusters was determined by using the factoextra (Kassambara and Mundt, 2016) package. The hierarchical cluster dendrogram was created by using denextend (Galili and Jefferis, 2014) and the circilize (Zuguang Gu, 2024) package.

Results

Quantitative data analysis

Descriptive statistics and Shannon-Weaver diversity indices for 12 quantitative traits are presented in Table 2. Considering the entire group of accessions, the coefficient of variation ranged from 6.24% to 22.92%, indicating a very low to high variability among quantitative traits (> 20%). Shannon–Weaver's diversity index ranges from 0.21–0.71, showing a very low to moderate level of diversity among the accessions for quantitative traits. Intra-accession variability was also observed, indicating that some accessions contained noticeable heterogeneity within the same accession.

Among the evaluated traits, days to first male flowering ranged from 72 to 91 days, with a mean of 78.64 days, with a very low coefficient of variation (CV) of 7.91% and a moderate Shannon-Weaver diversity index (H') of 0.57. Similarly, days to first female flowering spanned from 81 to 100 days, with an average of 89.09 days, and showed relatively very lower variability (CV = 6.24%) but had the highest diversity index among other variables (H' = 0.71). Regarding fruit characteristics, fruit length varied between 28.20cm and 34.50cm, averaging 31.50cm. The trait displayed very low variation (CV = 7.33%) and Shannon Weaver index (H' = 0.43). In contrast, fruit breadth ranged from 13.05cm to 18.37cm, with a mean value of 15.67cm. The CV for fruit breadth was 8.92%, accompanied by an H' value of 0.57. Total yield per hectare exhibited a wide range from 9.40t/ha to 16.07t/ha, with a mean yield of 11.43t/ha. Notably, this trait had a moderate (CV = 17.95%), reflecting considerable moderate yield variability among the accessions. However, the Shannon-Weaver index was the lowest (H' = 0.21) among traits.

Qualitative data analysis

The Shannon–Weaver diversity index (H′) among the cucumber accessions ranged from 0.44 to 0.99 (Figure 2), indicating low to very high diversity. Very high diversity (>0.80) was recorded for stem shape, stem colour, leaf size, leaf pubescence density, fruit surface, stem pubescence density, leaf glossiness, fruit shape, leaf lobes, leaf colour, fruit set, and immature fruit colour. Traits such as; fruit stem colour, leaf blade tip angle and blossom end fruit surface show the medium diversity (0.40-0.60). Only a single trait, spine colour showed low (0.10-0.40) diversity among the accessions.

Leaf brightness was mainly intermediate (45.45%), with fewer accessions showing dull or acute leaves. The leaf blade angle was largely acute (81.82%) while obtuse blades occurred less frequently. Leaf lobing was mostly intermediate (72.73%) and strongly lobed forms were comparatively rare. Leaf colour was predominantly light green (72.73%), with green types accounting for 27.27% (Figure 3a). Leaf pubescence density was generally intermediate (45.45%), while sparse and dense pubescence occurred less commonly. Leaf size was mainly large, followed by intermediate-sized leaves. Stem colour was largely uniform, being dominated by intermediate shades, with lighter colours also present. Stem pubescence density was most often intermediate, though dense pubescence was also common, and sparse pubescence occurred infrequently (Figure 3b). Stem shape was primarily angular (54.55%), though round stems were also well represented. Immature fruit colour was mainly green (63.64%), while light green and whitish green types each constituted 18.18%. Fruit surface texture was largely smooth, followed by ridged and rough surfaces. Spine colour was highly uniform, being dominated by whitish green spines, while other colours were rare (Figure 3c). Fruit size was mostly medium, with smaller fruits occurring less frequently.

Table 2. Descriptive statistics of the cucumber accessions. Min, minimum; Max, maximum; SE_m, standard error of mean; SD, standard deviation; CV, coefficient of variation; H', notation for Shannon-Weaver diversity index.

Variables	Min	Max	Mean± SEm	SD	CV (%)	H'
Days to first male flowering	72	91	78.64±1.87	6.22	7.91	0.57
Days to first female flowering	81	100	89.09±1.68	5.56	6.24	0.71
Length of petiole (cm)	9.03	16.63	12.85±0.76	2.52	19.64	0.67
Leaf length (cm)	12.20	17.87	14.48±0.50	1.66	11.44	0.50
Leaf width (cm)	14.17	23.60	17.50±0.77	2.55	14.59	0.29
Stem diameter (mm)	4.70	6.56	5.64±0.19	0.63	11.10	0.67
Distance between nodes (cm)	9.80	14.70	11.39±0.45	1.48	13.02	0.58
Length of peduncle (cm)	2.20	4.90	3.30±0.23	0.76	22.92	0.57
Peduncle thickness (mm)	1.88	3.98	2.87±0.18	0.60	20.95	0.61
Fruit length (cm)	28.20	34.50	31.50±0.70	2.31	7.33	0.43
Fruit breadth (cm)	13.05	18.37	15.67±0.42	1.40	8.92	0.57
Total yield (t/ha)	9.40	16.07	11.43±0.62	2.05	17.95	0.21

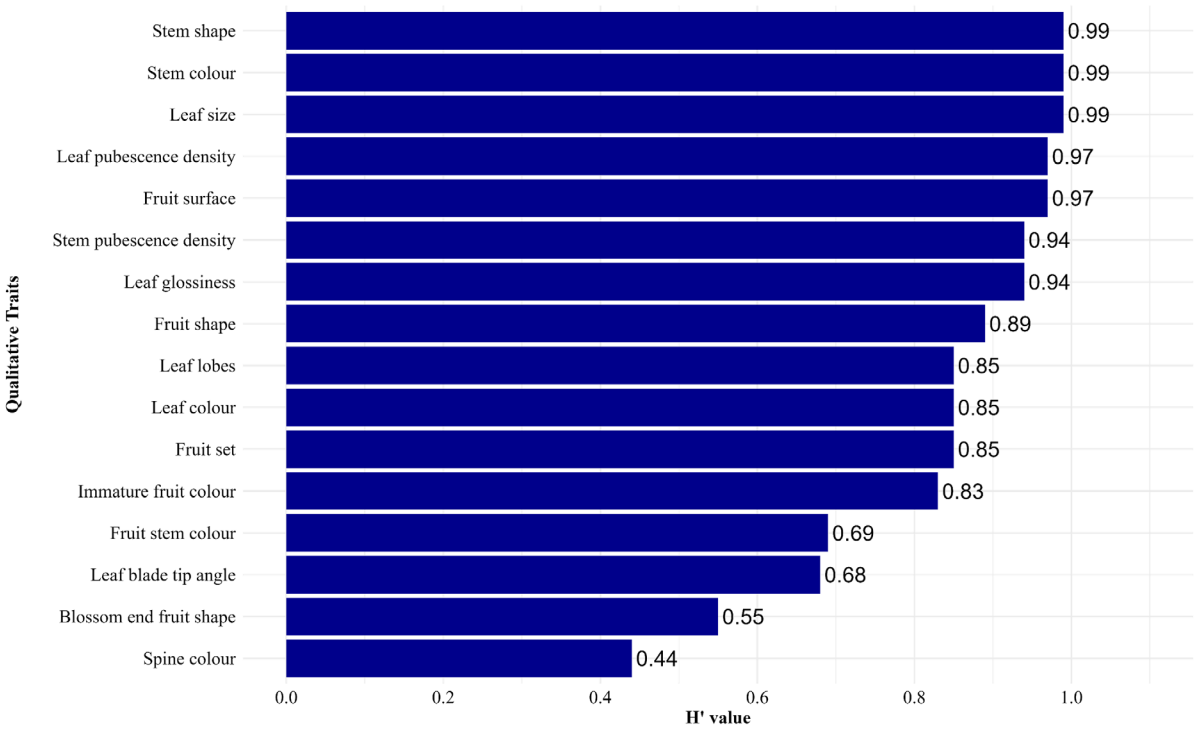


Figure 2. Shannon–Weaver Diversity Index of 16 qualitative traits of 11 Nepalese cucumber accessions

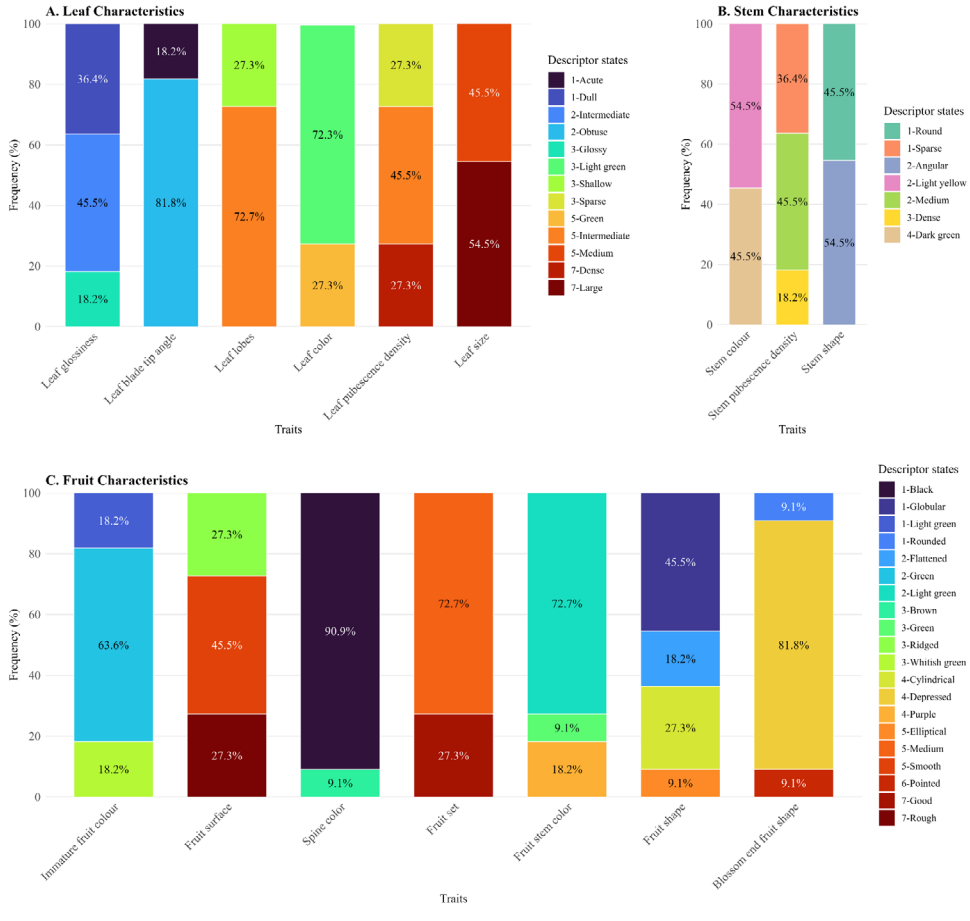


Figure 3. Variation observed among the qualitative morphological traits of cucumber accessions. Stacked bar charts showing frequency distribution of descriptor states. A, leaf characteristics; B, stem characteristics; C, fruit characteristics.

Fruit stem colour was predominantly light green, with green and intermediate shades observed in fewer accessions. Blossom-end fruit shape was largely depressed, with pointed and medium forms occurring only rarely (Figure 3c). Fruit shape was mainly elliptical (54.54%), followed by cylindrical (36.36%) and globular (9.09%) (Figure 4).

Principal component analysis

PCA identifies key traits contributing most to genetic variation and assists breeders in improving traits with low heritability, particularly in early generations (Golparvar *et al*, 2006; Ahmadizadeh and Felenji, 2011). Dimension-1 contributed 37.9% of the total variance, followed by dimension-2 with 17.5% variance, and cumulatively explained 55.48% of the total phenotypic variation (Figure 5). Traits such as leaf length (LW), leaf width (LW), length of petiole (LOP), distance between nodes (DBN), fruit length (FL), total yield (TY), fruit breadth (FB), stem diameter (SD), and peduncle thickness (PT) are positively associated with PC1, indicating that these traits contribute strongly to variation along this axis. Days to first male flowering (DTFMF) and days to first female flowering (DTFFF) load negatively on PC1, showing an inverse relationship with yield and growth-related traits. Length of peduncle (LPD) shows a strong negative loading on PC2, suggesting it contributes mainly to variation along the second axis. Accessions such as CO10293, CO13634 and CO14392 are positioned in the direction of yield and growth-related traits, indicating superior performance for these traits. CO14765 and CO13538 are associated with delayed flowering traits (DTFFF, DTFMF). Bhaktapur Local is closely associated with LPD, suggesting higher expression of this trait. SIRG-024-10CU lies far from most trait vectors, indicating distinct performance or divergence from other landraces.

Table 3 presents the eigenvalues, percentage of variance explained, cumulative variance and the coefficient (loading) vectors of each variable for the first three principal components (PCs). The first three PCs, cumulatively explaining ~70% of the total variation (PC1, 37.9%; PC2, 17.6%; PC3, 14.5%). PC1 was mainly shaped by traits related to plant size and productivity. Leaf width (0.41) and fruit breadth (0.41) contributed the most, along with leaf length (0.36), total yield (0.36) and petiole length (0.30). Overall, this component captured differences in leaf morphology and fruit size among the accessions. PC2 reflected variation more closely linked to flowering behaviour and plant structure. The distance between nodes (0.50) had the strongest influence, followed by days to first male flowering (0.43), days to first female flowering (0.39), petiole length (0.39) and peduncle length (0.35). These traits together highlight differences in how plants grow and transition to flowering. PC3 was strongly shaped by fruit length (0.48) and peduncle length (0.47). Total yield (0.32) also contributed to this component. This suggests that PC3 mainly differentiated accessions based on fruit elongation and key reproductive phenological traits. PCA indicated that traits such as leaf length, days to first female flowering, days to male first flowering, fruit breadth, length of peduncle, fruit length, total yield and leaf length had the highest loadings across the first three components, demonstrating their major contribution to the phenotypic variation observed among the Nepalese cucumber accessions.

Figure 4. Fruit shape diversity among the cucumber accessions

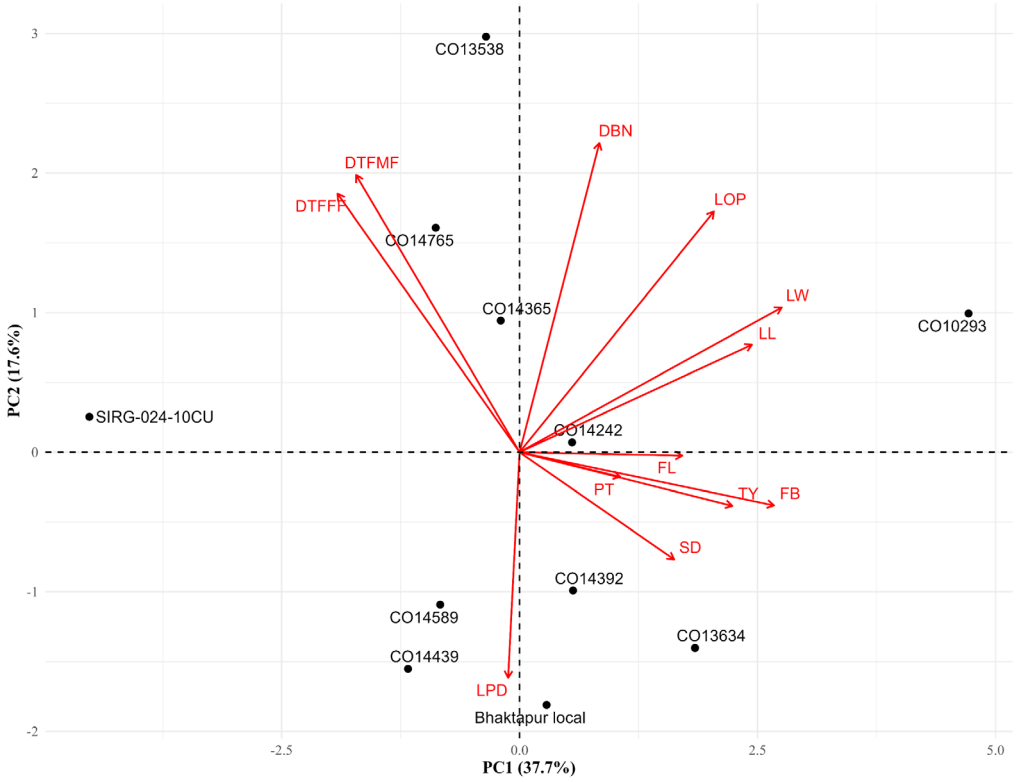


Figure 5. PCA-Biplot of Nepalese cucumber accessions. Red arrows represent trait vectors, and black points indicate specific accessions. The direction and length of each arrow reflect the correlation and relative contribution of each trait to the principal components, respectively. DTFMF: Days to first male flowering; DTFFF: Days to first female flowering; LOP: Length of petiole (cm); LL: Leaf length (cm); LW: Leaf width (cm); SD: Stem diameter (mm); DBN: Distance between nodes (cm); LPD: Length of peduncle (cm); PT: Peduncle thickness (mm); FL: Fruit length (cm); FB: Fruit breadth (cm), and TY: Total yield (t/ha).

Table 3. Contribution percentage and major characters associated with the first three principal components (PCs), of cucumber accessions, along with their eigenvectors.

Dimensions	**PC-1**	**PC-2**	**PC-3**
Eigenvalue	4.55	2.11	1.74
Variance (%)	37.93	17.55	14.49
Cumulative variance (%)	37.93	55.48	69.97
Variables		**Coefficient vectors**	
Days to first male flowering	-0.266	0.434	0.326
Days to first female flowering	-0.293	0.395	0.276
Length of petiole (cm)	0.306	0.392	-0.114
Leaf length (cm)	0.369	0.16	0.113
Leaf width (cm)	0.414	0.229	0.105
Stem diameter (mm)	0.244	-0.158	-0.262
Distance between nodes (cm)	0.122	0.508	-0.174
Length of peduncle (cm)	-0.019	-0.352	0.471
Peduncle thickness (mm)	0.16	-0.055	0.301
Fruit length (cm)	0.258	-0.015	0. 482
Fruit breadth (cm)	0.404	-0.084	-0.175
Total yield (t/ha)	0.335	-0.078	0.327

UPGMA cluster analysis

A dendrogram was constructed by using the unweighted pair group method with arithmetic mean (UPGMA) or average clustering method based on Euclidean distance across the 11 cucumber accessions (Figure 6). The average silhouette method identified a total of 4 clusters at 80% level of similarity. The cophenetic correlation approach identifies the best clustering, with higher values indicating a more accurate representation of original distances (Saraçli *et al*, 2013). The average method with a maximum cophenetic correlation (0.82) suggests an accurate depiction of the cluster. Cluster analysis grouped the accessions into 4 clusters based on 12 quantitative traits. Cluster I is represented by red colour, Cluster II by green, Cluster III by blue, whereas Cluster IV is represented by purple. Clusters I, II, III and IV consist of two, four, four and one accessions, respectively (Figure 6 and Table 4).

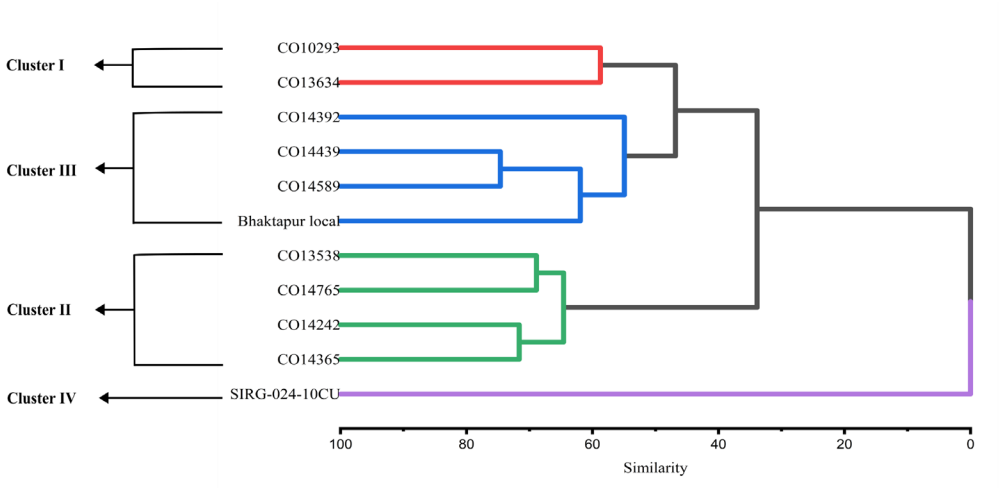


Figure 6. UPGMA cluster dendrogram of 11 cucumber accessions based on 12 quantitative traits using Euclidean distance

Table 4. Cluster characteristics of cucumber accessions evaluated for 12 quantitative traits. *, significant ($p \leq 0.05$); **, highly significant ($p \leq 0.01$); ***, very highly significant ($p \leq 0.001$); ns, not significant ($p > 0.05$).

Cluster	I	II	III	IV	Significance
Number of Accessions	2	4	4	1	
Days to first male flowering	75.50	82.67	72.83	91.33	**
Days to first female flowering	84.67	92.33	85.75	99.67	*
Length of petiole(cm)	15.49	14.05	11.27	9.03	**
Leaf length (cm)	15.80	14.94	13.94	12.20	ns
Leaf width (cm)	20.99	18.08	16.01	14.17	**
Stem diameter (mm)	6.54	5.23	5.67	5.38	ns
Distance between nodes (cm)	11.98	12.31	10.58	9.80	*
Length of peduncle (cm)	3.69	3.15	3.19	3.60	***
Peduncle thickness (mm)	3.09	2.84	2.87	2.53	ns
Fruit length (cm)	32.52	32.01	31.12	28.93	*
Fruit breadth (cm)	17.13	15.28	15.96	13.05	*
Total yield (t/ha)	14.98	10.97	10.64	9.40	*

Cluster analysis revealed clear and structured phenotypic differentiation among the cucumber accessions. Cluster I comprised two accessions (CO10293 and CO13634) and was distinguished by superior yield potential (14.98 t/ha; TY) coupled with vigorous vegetative growth, as reflected by significantly larger leaf width (LW). In contrast, Cluster II grouped four accessions (CO13538, CO14242, CO14365 and CO14765) and was characterized by significantly delayed flowering (DTFMF, DTFFF), moderate petiole length (LOP), intermediate stem diameter (SD ns), fruit length comparable to Cluster I (FL), and moderate yield performance (TY). Cluster III also consisted of four accessions (CO14439, CO14589, Bhaktapur Local and CO14392) and exhibited relatively earlier flowering behavior (DTFMF, DTFFF), moderate stem diameter (SD ns), and reduced fruit size (FL, FB), which was associated with a comparatively lower yield (TY) than Clusters I and II. Cluster IV contained a single accession (SIRG-024-10CU), which remained isolated and merged with other clusters at a very low similarity coefficient, indicating maximum phenotypic divergence (Table 5). This accession was characterized by markedly late flowering (DTFMF, DTFFF), significantly reduced vegetative traits (LOP, LW), and the lowest yield performance (TY).

Table 5. Cluster centroid distance

	Cluster1	Cluster2	Cluster3	Cluster4
Cluster1	0	8.918	11.974	25.377
Cluster2	8.918	0	12.555	23.731
Cluster3	11.974	12.555	0	14.158
Cluster4	25.377	23.7308	14.158	0

Phenotypic path analysis

Phenotypic path coefficient analysis was conducted to decompose the correlation coefficients into direct and indirect effects, aiming to determine the key traits influencing total yield. Total yield (TY) was used as the dependent variable to assess the direct and indirect contributions of the 11 associated traits, as illustrated in Figure 7. The peduncle thickness (0.522), leaf width (0.414), length of peduncle (0.408), length of petiole and stem diameter showed a high positive direct effect, suggesting that larger vegetative and fruit traits favour yield (Lenka and Misra, 1973). The fruit length (FL) exhibited a moderate positive direct effect (0.260) on yield, indicating that longer fruits contribute significantly to higher total yield. In contrast, fruit breadth (FB) and leaf length (LL) had negative direct effects (-0.275 and -0.365 respectively), implying that early flowering increases yield by extending the total duration of the reproductive phase. Interestingly, although fruit breadth (FB) and leaf length (LL) showed negative direct effects, their total phenotypic correlations with yield were positive, indicating that their overall positive influence is mediated through positive indirect effects on other yield components.

The model showed a high coefficient of determination (R^2 = 0.893) and a residual effect of 0.327, indicating that 89.3% of the variation in cucumber yield is explained by the traits included in the study. Overall, traits related to fruit size, fruit breadth, fruit length, leaf width, stem diameter, peduncle thickness, length of petiole, length of peduncle and distance between the nodes were key contributors to yield enhancement, whereas early flowering traits exerted minor negative influences.

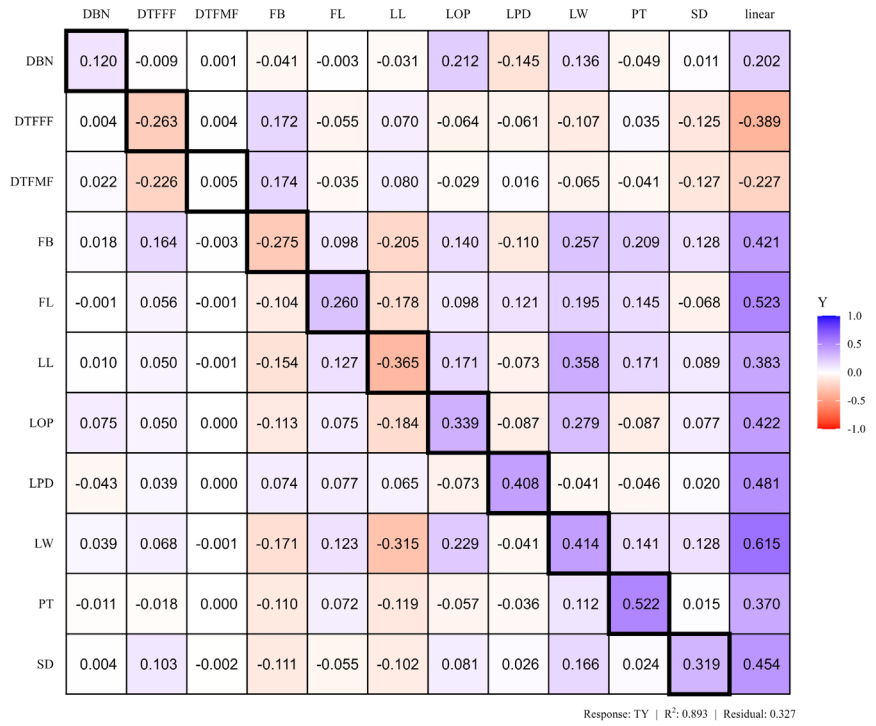


Figure 7. Heatmap showing the phenotypic path analysis of 11 quantitative traits on total yield. Diagonal values (bold boxes) represent direct effects and off-diagonal values represent indirect effects on total yield (TY). The linear column indicates total Pearson correlation. The color-coded scale indicates the nature of these relationships, with blue representing a positive effect and red representing a negative effect, while the color intensity reflects the magnitude of the impact. DTFMF, days to first male flowering; DTFFF, days to first female flowering; LOP, length of petiole (cm); LL, leaf length (cm); LW, leaf width (cm); SD, stem diameter (mm); DBN, distance between nodes (cm); LPD, length of peduncle (cm); PT, peduncle thickness (mm); FL, fruit length (cm); FB, fruit breadth (cm); TY, total yield (t/ha).

Genetic variability and heritability

GCV and PCV are considered ideal criteria for measuring variability (Yadav *et al*, 2021). The estimates of genetic variability parameters for 12 quantitative traits of cucumber revealed that phenotypic variance was higher than genotypic variance for all traits, indicating the influence of environmental factors on trait expression (Table 6). Moderate to high GCV and PCV were observed for length of peduncle, length of petiole, total yield, stem diameter and days to first male flowering, suggesting the presence of substantial variability among the accessions. Days to first male flowering, stem diameter and length of peduncle exhibited relatively high heritability (0.84, 0.70, and 0.73, respectively) along with moderate to high genetic advance as a percentage of mean, indicating the predominance of additive gene action and good prospects for improvement through direct selection. In contrast, traits such as leaf length, peduncle thickness, fruit length and fruit breadth showed low heritability and low genetic advance, reflecting strong environmental influence and limited response to selection. Total yield per hectare showed moderate genetic variability but low heritability, suggesting that yield is a complex trait influenced by multiple components and environmental conditions; therefore, indirect selection through yield-related traits with higher heritability would be more effective for genetic improvement in cucumber.

Discussion

The present study revealed a wide range of diversity in both qualitative and quantitative traits as evidenced by Shannon-Weaver diversity index (Eticha *et al*, 2006). Very high diversity indices of economically important qualitative parameters, such as fruit shape and fruit set, indicate the opportunities for selection during crop improvement (Perry and McIntosh, 1991). Immature fruit colour is an important marketing trait in Nepalese cucumber. In our study, most of the accessions (63.64%) had green fruit colour. Ahmed *et al* (2022) have also found a similar variation in fruit colour, with green dominating (65.05%) among seven different colours. The intermediate leaf glossiness (45.45%) and medium pubescence density (45.45%) were predominant among the accessions. Both traits function as defensive features that reduce insect pest incidence through lowered palatability and mechanical deterrence (War *et al*, 2012; Zhang *et al*, 2012). In our study, most accessions exhibited elliptical fruit shape, followed by cylindrical shape. A similar result was reported by Suma *et al* (2021) with predominantly elongate fruit shapes among Indian landraces, with only a few globular types. The elliptical–elongate form is highly preferred by consumers as it yields a greater number of slices per fruit, making it particularly suitable for salad preparation.

The evaluated accessions also showed wide variation in leaf traits. In this study, CO10293 exhibited the largest leaf dimensions. A larger leaf area generally enhances light interception and photosynthate production, contributing to higher yields (Verma *et al*, 2020; Mainali and Jyakhwa, 2023). Yield performance may also be influenced by other traits such as source–sink dynamics, fruiting efficiency and genetic potential of the genotype (Wang *et al*, 2018). A considerable variation was observed for days to female flowering (81–100 days), which is an important indicator of earliness in cucumber production (Shah *et al*, 2016). Bhaktapur Local (control variety) exhibited the earliest anthesis, which was followed by accession CO14439, consistent with male flowers emerging earlier than female to ensure pollen availability (Gaikwad *et al*, 2011; Pusphalatha *et al*, 2016; Valcárcel *et al*, 2018). In contrast, flowering was late in the SIRG-024-10CU accession. Such accession is useful for extending harvest periods and season-long market supply (Ranjan *et al*, 2019). Differences in flowering onset may reflect genetic factors (Pusphalatha *et al*, 2016; Ahirwar *et al*, 2017; Owino *et al*, 2020), hormonal balance, seed vigour and soil fertility (Shah *et al*, 2016). Environmental factors such as temperature, light intensity and photoperiod are also known to influence flowering in cucumber (Thiruvengadam and Chung, 2014; Lewandowska-Sabat *et al*, 2017). Fruit length was notably superior in accessions CO14392 and CO14365, whereas CO10293 exhibited the greatest fruit breadth. Fruit width is a key trait influencing both yield potential and fruit quality. The wide variation observed in fruit diameter among accessions

Table 6. Estimates of genetic parameters of cucumber accessions. GV, genotypic variance; PV, phenotypic variance; GCV, genotypic coefficient of variance; PCV, phenotypic coefficient of variance; H^2_{bs}, heritability; GAM, genetic advance as a percentage of mean.

Traits	GV	PV	GCV (%)	PCV (%)	H^2_{bs}	GAM (%)
Days to first male flowering	37.48	44.49	7.79	8.49	0.84	14.73
Days to first female flowering	24.92	38.56	5.59	6.96	0.64	9.26
Length of petiole (cm)	4.66	9.76	16.80	24.32	0.47	23.92
Leaf length (cm)	0.92	6.38	6.62	17.44	0.14	5.18
Leaf width (cm)	4.15	11.27	11.63	19.18	0.36	14.54
Stem diameter (mm)	0.34	0.48	10.44	12.31	0.70	18.25
Distance between nodes (cm)	1.43	3.74	10.50	16.97	0.38	13.38
Length of peduncle (cm)	0.52	0.69	21.65	25.25	0.73	38.25
Peduncle thickness (mm)	0.10	0.86	11.54	32.39	0.12	8.47
Fruit length (cm)	3.34	9.30	5.80	9.68	0.35	7.16
Fruit breadth (cm)	1.21	3.43	7.02	11.83	0.35	8.59
Total yield (t/ha)	2.48	7.83	13.55	24.48	0.30	15.46

likely reflects underlying genetic differences, variations in hormonal regulation, and overall plant vigour. Fruit size is a critical determinant of marketability, consumer preference and overall crop value (Chikh-Rouhou *et al*, 2019; Ilahy *et al*, 2020; Ni *et al*, 2024). Farmers commonly favour long-fruited accession due to their association with higher yield potential (Baniya *et al*, 2006). Excessively large fruits are prone to carpel separation, which adversely affects fruit quality and reduces consumer acceptance (Subedi *et al*, 2024). Variations in cucumber fruit size and shape are primarily influenced by fruit cell number and hormonal regulation during ovary and fruit development (Gillaspy *et al*, 1993; Tanksley, 2004; Liu *et al*, 2020). A strong positive relationship between fruit traits (length and breadth) and yield per plant has also been reported (Diouf *et al*, 2023; Mainali *et al*, 2023). Breeders should prioritize these key traits, along with increased fruit count and larger fruit diameter, to enhance yield potential (Arunkumar *et al*, 2011; Chikh-Rouhou *et al*, 2024).

PCA is a multivariate method that simplifies complex, interrelated quantitative data by converting it into a reduced set of uncorrelated principal components, thereby revealing patterns and relationships among variables and observations (Abdi and Williams, 2010). PCA revealed that the Nepalese cucumber landraces exhibited substantial phenotypic variation. Traits such as leaf length, leaf width, fruit breadth, fruit length, peduncle length, days to first male and female flowering, and total yield were the most contributing traits for developing variation. A cluster dendrogram displays how closely related different genotypes are, helping breeders identify genetically diverse groups for selection and hybridization (Mohammadi and Prasanna, 2003). Cluster analysis revealed distinct phenotypic diversity among cucumber landraces: Cluster I was characterized by high yield; Cluster II by late maturity, longer petioles and greater internode length; Cluster III by earliness, early flowering, smaller leaves and shorter petioles; and Cluster IV by very late flowering. These distinct traits highlight potential sources for breeding programmes targeting high yield, early maturity, or compact plant architecture.

Phenotypic path analysis identified leaf width, peduncle length and fruit length as major contributors to total yield, exhibiting strong positive effects. These results indicate that selection for wider and longer fruits, larger leaves, optimal internode spacing and thicker peduncles can effectively enhance yield. Consequently, breeding programmes should prioritize these traits when selecting parent lines for the development of high-yielding cucumber varieties. In our study, PCV exceeded GCV for all traits, suggesting the strong influence of environmental factors (Gaikwad *et al*, 2011; Chalbi *et al*, 2023), whereas the small differences between GCV and PCV indicate a greater influence of genotype on the expression of a particular trait. High heritability was found for traits like days to first male flowering, length of peduncle and stem diameter, which is due to the additive genetic effects (Gaikwad *et al*, 2011; Johnson *et al*, 1955). High broad-sense heritability (H2bs) and genetic advance (GAM) found for the days to first male flowering, stem diameter and peduncle length suggested being more effective for selection. Traits expressing both high heritability and GAM, along with yield-contributing traits such as total yield, plant height and leaf length, should be emphasized in cucumber improvement programmes (Veena *et al*, 2012; Pushpalatha *et al*, 2016).

Conclusion

The Nepalese cucumber accessions exhibited substantial agromorphological variation, underscoring their potential for genetic improvement. Subsistence and commercial farmers in Nepal place greater emphasis on traits such as early flowering, higher fruit number and longer and wider fruits for market acceptance. Accordingly, breeders should prioritize accessions such as CO10293, CO14392 and CO13634 for their superior fruit dimensions and yield potential. Multivariate analysis highlighted key traits and superior parents for hybridization, with the observed phenotypic diversity providing valuable opportunities for yield improvement, selection and conservation. To fully exploit this variability, multi-location and multi-year evaluations, complemented by molecular characterization, are recommended.

Supplemental data

Supplemental Table 1. Morphological qualitative traits observed and their descriptive states

Supplemental Table 2. Morphological quantitative traits observed with descriptive states

Supplemental Table 3. Raw data of 12 quantitative traits recorded across replications of cucumber landraces

Acknowledgements

We would like to thank the National Agriculture Genetic Resources Centre, Genebank, Khumaltar, Lalitpur, for providing all the support for conducting the research.

Author contributions

Pradip Thapa was responsible for conceptualization, investigation, methodology and writing the original draft. Sandip Bohara contributed to data curation, formal analysis, software development and reviewing and editing the manuscript. Subechha Giri, Naturally K.C., and Basanta Kumar Rimal were involved in methodology and writing the original draft. Dr Bal Krishna Joshi oversaw project administration, supervision and validation.

Conflict of interest statement

We declare that there is no conflict of interest among the authors regarding this research work.

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

Abdi, H. and Williams, L. J. (2010). Principal component analysis. *Wiley interdisciplinary reviews: computational statistics* 2(4), 433-459. https://doi.org/10.1002/wics.101

Ahirwar, C. S., Singh, D. K., and Kushwaha, M. L. (2017). Assessment of genetic divergence in cucumber (*Cucumis sativus* L.) germplasm through clustering and principal component analysis. *Journal of Pharmacognosy and Phytochemistry*, 6(5), 804-807.

Ahmadizadeh, M., and Felenji, H. (2011). Evaluating diversity among potato cultivars using agro-morphological and yield

components in fall cultivation of Jiroft area. *American-Eurasian Journal of Agricultural & Environmental Sciences* 11(5), 655-662.

Ahmed, I., Rohman, M. M., Hossain, M. A., Molla, M. R., Azam, M. G., Hasan, M. M., ... and Hossain, A. (2022). A study on the phenotypic variation of 103 cucumber (*Cucumis sativus* L.) landraces for the development of desirable cultivars suitable for the changing climate. *Life* 12(8), 1235. https://doi.org/10.3390/life12081235

Arunkumar, K. H., Ramanjinappa, V., and Hugar, A. (2011). Association of yield and yield components in F_2 population of cucumber (*Cucumis sativus* L.). *Plant Archives* 11(1), 457-459.

Auguie, B., Antonov, A., and Auguie, M. B. (2017). Package 'gridExtra'. Miscellaneous functions for "grid" graphics, 9. https://doi.org/10.32614/CRAN.package.gridExtra

Azam, M. G., Sarker, U., Hossain, M. A., Mahabubul Alam, A. K. M., Islam, M. R., Hossain, N., and Alamri, S. (2024). Phenotypic diversity in qualitative and quantitative traits for selection of high yield potential field pea genotypes. *Scientific Reports* 14(1), 18561. https://doi.org/10.1038/s41598-024-69448-7

Baniya, B., Joshi, B., Barjracharya, J., Yadav, R., Tiwari, R., Shrestha, P., Gauchan, D., and Shrestha, P. (2006). On-farm diversity management of sponge gourd, cucumber and buckwheat in Nepal. In *On-farm Management of agricultural biodiversity in Nepal: Lessons learned,* eds. B. Sthapit and D. Gauchan (Proceedings of national symposium), 64-73. http://www.plantarchives.org/

Bisht, I. S., Bhat, K. V., Tanwar, S. P. S., Bhandari, D. C., Joshi, K., & Sharma, A. K. (2004). Distribution and genetic diversity of *Cucumis sativus* var. *hardwickii* (Royle) Alef in India. *The Journal of Horticultural Science and Biotechnology* 79(5), 783-791. https://dx.doi.org/10.1080/14620316.2004.11511843

Cebeci, E., Gozen, V., Keskin, L., and Yildirim, A. (2020). Morphologic and molecular assessments of cucumber (*Cucumis sativus* L.) landraces. *Notulae Botanicae Horti Agrobotanici Cluj-Napoca* 48(2), 604-614. https://doi.org/10.15835/nbha48211932

Chakraborty, S., and Rayalu, S. (2021). Health beneficial effects of cucumber. In *Cucumber Economic Values and its Cultivation and Breeding,* ed. H. Wang (Intech Open), 99-107. https://doi.org/10.5772/intechopen.87508

Chalbi, A., Chikh-Rouhou, H., Mezghani, N., Slim, A., Fayos, O., Bel-Kadhi, M. S., Garcés-Claver, A. (2023). Genetic Diversity Analysis of Onion (*Allium cepa* L.) from the Arid Region of Tunisia Using Phenotypic Traits and SSR Markers. *Horticulturae* 9, 1098. doi: https://doi.org/10.3390/horticulturae9101098

Cheng, J., Wen, S., and Bie, Z. (2019). Overexpression of hexose transporter CsHT3 increases cellulose content in cucumber fruit peduncle. *Plant Physiology and Biochemistry* 145, 107-113. https://doi.org/10.1016/j.plaphy.2019.10.034

Chikh-Rouhou, H., Abdedayem, W., Solmaz, I., Sari, N., Garcés-Claver, A. (2023). Melon (*Cucumis melo* L.): Genomics and Breeding. In: Singh, S., Sharma, D., Sharma, S.K., Singh, R. (eds.) Smart Plant Breeding for Vegetable Crops in Post-Genomics Era. pp 25-52. Springer, Singapore. https://doi.org/10.1007/978-981-19-5367-5_2

Chikh-Rouhou, H., Tlili, I., Ilahy, R., R'Him, T., Sta-Baba, R. (2019). Fruit quality assessment and characterization of melon genotypes. *Int. J. Veg. Sci.* 27, 3–19. doi: https://doi.org/10.1080/19315260.2019.1692268

Chikh-Rouhou, H., Kienbaum, L., Gharib, A., Fayos, O., Garcés-Claver, A. (2024). Combining ability and hybrid breeding in Tunisian melon (*Cucumis melo* L.) for fruit traits. *Horticulturae,* 10 (7), 724. https://doi.org/10.3390/horticulturae10070724

Comtois, D. and Comtois, M.D. (2016). Packages 'summary tools'. https://doi.org/10.32614/CRAN.package.summarytools

Diouf, S., & Sambou, A. (2023). Agro-morphological characterization of four varieties of cucumber from *Cucumis sativus* L. and *Cucumis metuliferus* E. Mey. Ex Naudin in Senegal. *Journal of Horticulture and Postharvest Research* 6(2), 131-144.

Eticha, F., Belay, G., & Bekele, E. (2006). Species Diversity in Wheat Landrace Populations from two Regions of Ethiopia. *Genetic Resources and Crop Evolution*, *53*(2), 387–393. https://doi.org/10.1007/s10722-004-6095-z

FAOSTAT. (2024). Food and Agriculture Organization of the United Nations. Crop. Prod. Data 2023. Available online: http://www.fao.org/faostat/en/#data (accessed on 8 March 2025).

Friedman, J. M., Hunt Jr, E. R., and Mutters, R. G. (2016). Assessment of leaf color chart observations for estimating maize chlorophyll content by analysis of digital photographs. *Agronomy Journal* 108(2), 822-829. DOI: https://doi.org/10.2134/agronj2015.0258

Fu, X., Zhou, L., Huang, J., Mo, W., Zhang, J., Li, J., ... and Huang, X. (2013). Relating photosynthetic performance to leaf greenness in litchi: A comparison among genotypes. *Scientia Horticulturae* 152, 16-25. https://doi.org/10.1016/j.scienta.2013.01.001

Gaikwad, A. G., Musmade, A. M., Dhumal, S. S., and Sonawane, H. G. (2011). Variability studies in cucumber (*Cucumis sativus* L.). *Ecology, Environment and Conservation* 17(4), 799-802.

Galili, T., and Jefferis, G. (2014). dendextend: Extending "dendrogram" Functionality in R (p. 1.19.0) [Dataset]. https://doi.org/10.32614/CRAN.package.dendextend

Gao, Y., Islam, M. S., Tian, J., Lui, V. W. Y., and Xiao, D. (2014). Inactivation of ATP citrate lyase by Cucurbitacin B: A bioactive compound from cucumber, inhibits prostate cancer growth. *Cancer Letters* 349(1), 15-25. https://doi.org/10.1016/j.canlet.2014.03.015

Gautam, I. P., Pradhan, N. G., Subedi, S., and Thakur, M. K. (2021). Evaluation of cucumber hybrids for yield and quality under plastic house and open field conditions. *Nepalese Horticulture* 15, 52-63. https://doi.org/10.3126/nh.v15i0.36649

Genebank. (2023). *Annual Report 2079/80 (2022/23)*. National Agriculture Genetic Resources Centre, NARC, Khumaltar, Lalitpur, Nepal.

Genebank. (2024). *Annual Report 2080/81 (2023/24)*. National Agriculture Genetic Resources Centre, NARC, Khumaltar, Lalitpur, Nepal.

Ghimire, K. H., and Magar, R. T. (2017). Diversity in Nepalese wheat genetic resources as revealed by agro-morphological markers. *International Journal of Scientific and Engineering Research* 8(7), 1646-1651.

Gillaspy, G., Ben-David, H., and Gruissem, W. (1993). Fruits: A developmental perspective. *The Plant Cell* 5(10), 1439.

Golparvar, A.R., Pirbalouti, A.G., and Madani, H. (2006): Genetic control of some physiological attributes in wheat under drought stress conditions. *Pak. J. Bio. Sci.*, 9(8): 1442-1446.

Grosjean, P., Ibanez, F., Etienne, M., and Grosjean, M. P. (2018). Package 'pastecs'. Pastecs: Package for Analysis of Space-Time Ecological Series. Url: https://cran. r-project.

org/web/packages/pastecs/pastecs. doi: https://doi.org/10.32614/CRAN.package.pastecs

Husson, F., Josse, J., Le, S., Mazet, J., & Husson, M. F. (2016). Package 'factominer'. *An R package*, *96*(96), 698. http://factominer.free.fr

Ilahy, R., Tlili, I., Rouhou, H., Siddiqui, M., Mishra, P., Kuchi, V., Homa, F., Hdider, C., Jebari, H., Lenucci, M. S. (2020). Determining the main agronomic traits of snake melon (Cucumis melo var. flexuosus L.) fruits as affected by genotypic differences. *Acta Hortic. Solis Culin.* 34, 113–119. doi: https://doi.org/10.13128/ahsc-8254

Johnson, H. W., Robinson, H. F., and Comstock, R. E. (1955). Estimates of genetic and environmental variability in soybeans. *Agronomy Journal* 47, 314-318.

Kassambara, A. and Mundt, F. (2017). Package 'factoextra'. Extract and visualize the results of multivariate data analyses,76(2). https://doi.org/10.32614/CRAN.package.factoextra

Khulakpam, N. S., Singh, V., and Rana, D. K. (2015). Medicinal importance of cucurbitaceous crops. *International Research Journal of Biological Sciences* 4(6), 1-3.

Kirkbride, J. H. (1993). *Biosystematic monograph of the genus Cucumis (Cucurbitaceae)*. Parkway Publishers, Boone, North Carolina.

Lenka, D. and Misra, B. (1973) Path Coefficient Analysis of Yield in Rice Varieties. Indian Journal of Agricultural Science, 43, 376-379.

Lewandowska-Sabat, A. M., Fjellheim, S., Olsen, J. E., and Rognli, O. A. (2017). Local populations of Arabidopsis thaliana show clear relationship between photoperiodic sensitivity of flowering time and altitude. *Frontiers in Plant Science* 8, 1046. https://doi.org/:10.3389/fpls.2017.01046

Liu, X., Pan, Y., Liu, C., Ding, Y., Wang, X., Cheng, Z., and Meng, H. (2020). Cucumber fruit size and shape variations explored from the aspects of morphology, histology, and endogenous hormones. *Plants* 9(6), 772. https://doi.org/10.3390/plants9060772

Lv, J., Qi, J., Shi, Q., Shen, D., Zhang, S., Shao, G., ... and Huang, S. (2012). Genetic diversity and population structure of cucumber (*Cucumis sativus* L.). *PLoS ONE* 7(10), e46919. https://doi.org/10.1371/journal.pone.0046919

Mainali, R. P., and Jyakhwa, S. (2023). Morpho-agronomic characterization of cucumber (*Cucumis sativus*) germplasm for yield and yield-associated traits. *International Journal of Applied Sciences and Biotechnology* 11(1), 30-36. https://doi.org/10.3126/ijasbt.v11i1.53702

MoALD. (2025). *Statistical information on Nepalese agriculture*. Ministry of Agriculture and Livestock Development (2022/23). Singha Durbar, Kathmandu: MoALD. Retrieved from https://moald.gov.np/

Mohammadi, S. A., and Prasanna, B. M. (2003). Analysis of genetic diversity in crop plants—salient statistical tools and considerations. *Crop science* 43(4), 1235-1248.

Naegele, R.P., and Wehner, T.C. (2016). Genetic resources of cucumber. In *Genetics and genomics of Cucurbitaceae*, eds. R. Grumet, N. Katzir, and J. Garcia-Mas (Cham: Springer), 61–86. https://doi.org/10.1007/7397_2016_15

Ni, F., Li, Z., and Huang, J. (2024). Worldwide productivity and research trend on fruit quality: a bibliometric study. *Frontiers in Plant Science* 14, 1-8. https://doi.org/10.3389/fpls.2023.1294989

Niyi, O. H., Jonathan, A. A., and Ibukun, A. O. (2019). Comparative assessment of the proximate, mineral composition and mineral safety index of peel, pulp and seeds of Cucumber (*Cucumis sativus*). *Open Journal of Applied Sciences* 9(9), 691. https://doi.org/10.4236/ojapps.2019.99056

Olivoto, T., Lúcio, A. D. (2020). metan: An R package for multi-environment trial analysis. *Methods Ecol. Evol.* 11, 783–789. doi: https://doi.org/10.1111/2041-210X.13384

Owino, M. H., Gichimu, B. M., and Muturi, P. W. (2020). Agro-morphological characterization of horned melon ('*Cucumis metuliferus*') accessions from selected agro-ecological zones in Kenya. *Australian Journal of Crop Science* 14(9), 1487-1496. https://doi.org/10.21475/ajcs.20.14.09.p2642

Perry, M. C., and McIntosh, M. S. (1991). Geographical patterns of variation in the USDA soybean germplasm collection: I. Morphological traits. *Crop Science* 31(5), 1350-1355. https://doi.org/10.2135/cropsci1991.0011183X003100050054x

Popat, R., Patel, R., and Parmer, D. (2020). Variability: genetic variability analysis for plant breeding research https://CRAN.R-project.org/package=variability.

Pushpalatha, N., Anjanappa, M., Devappa, V., and Pitchaimuthu, M. (2016). Genetic variability and heritability for growth and yield in cucumber (*Cucumis sativus* L.). *Journal of Horticultural Sciences* 11(1), 33-36. https://doi.org/10.24154/jhs.v11i1.100

R Core Team (2025). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Ranjan, P., Gangopadhyay, K. K., Bag, M. K., Roy, A., Srivastava, R., Bhardwaj, R., and Dutta, M. (2015). Evaluation of cucumber (*Cucumis sativus* L.) germplasm for agronomic traits and disease resistance and estimation of genetic variability. *Indian Journal of Agricultural Sciences* 85(2), 234-239. http://epubs.icar.org.in/ejournal/index.php/IJAgS/article/view/46516/20091

Ranjan, P., Pandey, A., Munshi, A. D., Bhardwaj, R., Gangopadhyay, K. K., Malav, P. K., ... and Kumar, A. (2019). Orange-fleshed cucumber (*Cucumis sativus* var. *sativus* L.) germplasm from North-East India: agro-morphological, biochemical and evolutionary studies. *Genetic Resources and Crop Evolution* 66(6), 1217-1230. https://doi.org/10.1007/s10722-019-00778-6

Raza, A., Ayyub, C. M., Ghani, M. A., and Ahmed, N. (2020). Assessment of morphological diversity among indigenous cucumber germplasm of Pakistan. *Pakistan Journal of Agricultural Sciences* 57(6), 1573-1580. https://doi.org/10.21162/PAKJAS/20.9956

Renner, S. S., Schaefer, H., and Kocyan, A. (2007). Phylogenetics of *Cucumis* (Cucurbitaceae): Cucumber (*C. sativus*) belongs in an Asian/Australian clade far from melon (*C. melo*). *BMC Evolutionary Biology* 7, 1-11. https://doi.org/10.1186/1471-2148-7-58

Saracli, S., Dogan, N., and Dogan, I. (2013). Comparison of hierarchical cluster analysis methods by cophenetic correlation. *Journal of Inequalities and Applications* 2013:203. https://doi.org/10.1186/1029-242X-2013-203

Sebastian, P., Schaefer, H., Telford, I. R., and Renner, S. S. (2010). Cucumber (*Cucumis sativus*) and melon (*C. melo*) have numerous wild relatives in Asia and Australia, and the sister species of melon is from Australia. *Proceedings of the National Academy of Sciences* 107(32), 14269-14273. https://doi.org/10.1073/pnas.1005338107

Seed Quality Control Centre (SQCC). (2024). *Notified or denotified varieties updated 2081-07-23*. Government of Nepal, Ministry of Agriculture and Livestock Development. Retrieved from https://www.sqcc.gov.np/storage/listies/November2024/notified%20and%20denotified%20varieties%20updated%202081-7-23.pdf

Shah, K. N., Rana, D. K., & Singh, V. (2016). Evaluation of different cucumber strain for various horticultural traits

under valley condition of Garhwal Himalaya. *J. Plant Dev. Sci,* 8(12), 599-603. url: https://jpds.co.in/wp-content/uploads/2025/11/05.-Naseer-Ahmed-Correct-Paper-1025.pdf

Shakya, S., Bhattarai, S., Tripathi, K., Sharma, M. (2006). Screening of cucumber germplasms for high temperature and long day photoperiodic tolerance. *J. Inst. Agric. Anim. Sci.* 27, 45–51. doi: https://doi.org/10.3126/jiaas.v27i0.694

Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal* 27(3), 379-423.https://doi.org/10.1002/j.1538-7305.1948.tb01338.x

Sharma, V., Sharma, L., and Sandhu, K. S. (2020). Cucumber (*Cucumis sativus* L.). In *Antioxidants in vegetables and nuts-properties and health benefits*, eds. G.A. Nayik and A. Gull (Singapore: Springer), 333–340. https://doi.org/10.1007/978-981-15-7470-2_17

Staub, J. E., Serquen, F. C., and McCreight, J. D. (1997). Genetic diversity in cucumber (*Cucumis sativus* L.): III. An evaluation of Indian germplasm. *Genetic Resources and Crop Evolution* 44, 315-326. https://doi.org/10.1023/A:1008639103328

Subedi, S., Bhandari, N., Basnet, M., Pradhan, N. G., Gautam, I. P. (2024). Evaluation of cucumber genotypes under plastic house and open field conditions in Lalitpur, Nepal. *Cogent Food Agric.* 10. doi: https://doi.org/10.1080/23311932.2023.2298531

Suma, A., Elsy, C. R., John, K. J., Pradeepkumar, T., Francies, R., Joseph, J., Santhoshkumar, A.V., Krishnan, S., and Yadav, S. K. (2021). Genetic diversity in Indian landraces of cucumber (*Cucumis sativus* L.) based on morpho-horticultural traits. *Indian Journal of Plant Genetic Resources* 34(3), 411-423. https://doi.org/10.5958/0976-1926.2021.00034.6

Tanksley, S. D. (2004). The genetic, developmental, and molecular bases of fruit size and shape variation in tomato. *The Plant Cell* 16, 181-189. https://doi.org/10.1105/tpc.018119

Thiruvengadam, M., and Chung, I. M. (2014). Optimization of factors influencing in vitro flowering of gherkin (*Cucumis anguria* L.). *Acta Biologica Hungarica* 65, 72-84. https://doi.org/10.1556/ABiol.65.2014.1.7

UPOV. (2007). Guidelines for the Conduct of Test for Distinctness, Uniformity and Stability for Cucumber, Gherkin (*Cucumis sativus* L.); International Union for the Protection of New Varieties of Plants: Geneva, Switzerland

Valcárcel, J. V., Peiró, R. M., Pérez-de-Castro, A., and Díez, M. J. (2018). Morphological characterization of the cucumber (*Cucumis sativus* L.) collection of the COMAV's Genebank. *Genetic Resources and Crop Evolution* 65(4), 1293-1306. https://doi.org/10.1007/s10722-018-0614-9

Venna, R., Sindhu, A.S., Pitchaimuthuverma, M., and Souravi, K. (2012). Genetic evaluation of Cucumber [*Cucumis sativus* L.] genotypes for some yield and related traits. *Electronic Journal of Plant Breeding* 3(3), 945-948. https://doi.org/10.37992/2023.1401.038

Verma, N., Kumar, R., Kaur, J., and Singh, D. (2020). Effect of different mode of pollination on fruit and seed characteristics of cucumber (*Cucumis sativus* L.). *Vegetable Science* 47(02), 248-253.

Villa, T. C. C., Maxted, N., Scholten, M., and Ford-Lloyd, B. (2005). Defining and identifying crop landraces. *Plant Genetic Resources* 3(3), 373-384. https://doi.org/10.1079/PGR200591

Wang, X., Bao, K., Reddy, U. K., Bai, Y., Hammar, S. A., Jiao, C., ... and Fei, Z. (2018). The USDA cucumber (*Cucumis sativus* L.) collection: genetic diversity, population structure, genome-wide association studies, and core collection development. *Horticulture research* 5. https://doi.org/10.1038/s41438-018-0080-8

War, A. R., Paulraj, M. G., Ahmad, T., Buhroo, A. A., Hussain, B., Ignacimuthu, S., and Sharma, H. C. (2012). Mechanisms of plant defense against insect herbivores. *Plant Signaling & Behavior* 7(10), 1306–1320. https://doi.org/10.4161/psb.21663

Weng, Y. (2021). Cucumis sativus chromosome evolution, domestication, and genetic diversity: Implications for cucumber breeding. In *Plant Breeding Reviews,* ed. I. Goldman (John Wiley & Sons), 9-111. https://doi.org/10.1002/9781119717003.ch4

Whitaker, T. W., and Davis, G. N. (1962). *Cucurbits: Botany, Cultivation, and Utilization.* Interscience Publishers.

Xiaolei, S., and Zhifeng, W. (2002). The optimal leaf area index for cucumber photosynthesis and production in plastic greenhouse. In *XXVI International Horticultural Congress: Protected Cultivation 2002: In Search of Structures, Systems and Plant Materials for 633,* ed. A.P. Papadopoulos (Interscience Publishers), 161-165.

Yadav, S., Singh, D. K., Singh, S. S., Bisht, Y. S., and Bhatt, R. (2021). Assessment of genetic variability in cucumber (*Cucumis sativus* L.). *IJCS* 9(1), 3347-3349. https://doi.org/10.22271/chemi.2021.v9.i1au.11753

Yang, L., Koo, D. H., Li, Y., Zhang, X., Luan, F., Havey, M. J., ... and Weng, Y. (2012). Chromosome rearrangements during domestication of cucumber as revealed by high□density genetic mapping and draft genome assembly. *The Plant Journal* 71(6), 895-906. https://doi.org/10.1111/j.1365-313X.2012.05017.x

Younuskunju, S., Mohamoud, Y. A., Mathew, L. S., Mayer, K. F., Suhre, K., and Malek, J. A. (2025). Genomic analysis of date palm fruit size traits and identification of candidate genes through GWAS. arXiv preprint arXiv:2501.11831. https://doi.org/10.32220/agriRxiv.2025.00299

Zeven, A. C. (1998). Landraces: A review of definitions and classifications. *Euphytica* 104, 127-139. https://doi.org/10.1023/A:1018683119237

Zhang, C., Pratap, A. S., Natarajan, S., Pugalendhi, L., Kikuchi, S., Sassa, H., ... and Koba, T. (2012). Evaluation of morphological and molecular diversity among South Asian germplasms of *Cucumis sativus* and *Cucumis melo*. *International Scholarly Research Notices* 2012(1), 134134. https://doi.org/10.5402/2012/134134

Zuguang Gu, M. (2024). Package "circlize" Title Circular Visualization. https://doi.org/10.1093/bioinformatics/btu393

Genetic Resources

ORIGINAL ARTICLE

Genetic Resources (2026), 7 (13), 103–117
DOI: 10.46265/genresj.TXWX7641
https://www.genresj.org
ISSN: 2708-3764

Characterizing genetic diversity within and between native Nordic horse breeds utilizing and comparing the EquCab3.0 and EquCab_Finn reference genomes

Nathalie Almaas Smogeli[a], Iryna Shutava[b], Signa Kallsoy Ravnafoss[c], Maria Kjetså[d], Juha Kantanen[e], Kisun Pokharel[e], Therese Selle[f], Sofia Mikko[b], Susanne Eriksson[b,*], Peer Berg[a]

[a]*Norwegian University of Life Sciences, Department of Animal and Aquacultural Sciences, P.O. Box 5003, NO-1432 Ås, Norway*
[b]*Swedish University of Agricultural Sciences, Department of Animal Biosciences, P.O. Box 7023, SE-75007 Uppsala, Sweden*
[c]*University of Copenhagen, Department of Biology, Nørregade 10, PO Box 2177, 1017, Copenhagen K, Denmark*
[d]*Nordic Genetic Resource Center (NordGen), Växhusvägen 12, 23456-Alnarp, Sweden*
[e]*Natural Resources Institute Finland, Tietotie 4, FI-31600 Jokioinen, Finland*
[f]*Norwegian Horse Association, Starumsvegen 71, 2850 Lena, Norway*

Abstract: Sustainable breeding of native breeds is essential to preserve genetic diversity and cultural heritage. Several native Nordic horse breeds are at risk of extinction and lack genetic characterization. This study aimed to analyze genetic variation and kinship within and among native Nordic horse breeds using whole-genome sequence data, and to compare results from using a Finnhorse genome assembly to that of the EquCab3.0 (Thoroughbred) reference genome. The breeds Dola Horse, North Swedish Horse and Coldblooded Trotter showed close genetic relationship for fixation index (0.03–0.08), and in principal component analysis. The other breeds showed stronger genetic differentiation, especially the Faroese Horse, with fixation index above 0.16 to all other breeds. This breed had the highest genomic inbreeding of 33% and the lowest heterozygosity of 12%. The Swedish Ardennes showed the lowest inbreeding at 14% and the highest heterozygosity of 16%. The mean identity by descent varied from 17% for Swedish Ardennes to 40% for Faroese Horses. The choice of reference genomes gave minor to moderate differences, suggesting that a closer related reference improves precision for fine mapping and understanding of the genetic landscapes of Nordic breeds. Together, the different analyses showed low genetic diversity in all breeds, and the general pattern of relatedness largely agreed with the known breed history. The results underline the importance of maintaining genetic diversity for the survival of the breeds.

Keywords: Whole genome sequence data, population genetics, genomic inbreeding, breed conservation, local breeds

Citation: Smogeli, N. A., Shutava, I., Ravnafoss, S. K., Kjetså, M., Kantanen, J., Pokharel, K., Selle, T., Mikko, S., Eriksson, S. and Berg, P. (2026) "Characterizing genetic diversity within and between native Nordic horse breeds utilizing and comparing the EquCab3.0 and EquCab_Finn reference genomes", *Genetic Resources*, 7(13), pp. 103–117. doi: 10.46265/genresj.TXWX7641.

Background

Preservation and understanding of genetic diversity are paramount for sustainable breeding practices, and to ensure the continued overall resilience of animals (FAO, 2022). With its climatic and geographical variations, the Nordic region has given rise to several horse breeds with unique characteristics and historical significance. These breeds often represent small, isolated populations that survived population bottlenecks, some of which share a common ancestry. Unfortunately, this has left several of the breeds endangered.

In this study, we focused on Swedish and Norwegian native horse breeds, with the addition of the Faroese Horse. While the Faroese Horse is critically endangered, breeds with endangerment of extinction include Swedish Ardennes, Dola Horse, Nordland/Lyngen Horse and Norwegian Fjord Horse. Vulnerable breeds include Coldblooded Trotter, Gotland Pony and North Swedish Horse (FAO, 2024; White *et al*, 2024).

The population history and current state vary between the Norwegian and Swedish native breeds. The first Swedish Ardennes (Svensk Ardenner) were imported to Sweden from

* Corresponding author: Susanne Eriksson (susanne.eriksson@slu.se)

Received: 12.08.2025 | Accepted: 20.02.2026 | Published online: 18.03.2026

the Ardennes in Belgium in 1873. The breed became very popular, especially in southern Sweden. Currently, there are 4,100 registered horses in the breed (Hästnäringens Nationella Stiftelse, 2021). In 2024, there were only 231 Swedish Ardennes mares used in breeding, which is less than 1% of the over 30,000 coverings by Swedish Ardennes stallions reported in 1945 (Carlström *et al*, 1946; Svenska Hästavelsförbundet, 2024).

The Dola Horse (Dølahest) was established as a breed in the mid-19th century because of a need for a horse for agriculture, forestry and transport (White *et al*, 2024). Today, there are approximately 3,900 Dola Horses, and 290 coverings were done in 2023 (Norsk Hestesenter, 2023). It is permitted to cover type-approved Coldblooded Trotter mares with approved Dola Horse stallions for offspring to be included in the Dola Horse studbook (Landslaget for Dølahest, 2024). Additionally, North Swedish Horse stallions can be used on Dola Horse mares with a quota if approved (Norsk Hestesenter, 2023).

North Swedish Horse (Nordsvensk Brukshäst) traces back to native horses present in prehistoric Sweden, which later had some influence from foreign breeds. This breed played a crucial role, particularly during the 19th century, for forestry and farming (Föreningen Nordsvenska Hästen, 2019; White *et al*, 2024). Efforts to preserve this horse type began at the end of the 19th century, and Norwegian stallions were introduced to strengthen the genetic pool. The breed association was formed in 1924. In 2020, the breed had about 5,100 horses (Hästnäringens Nationella Stiftelse, 2021), and 368 coverings were reported in Sweden 2024 (Svenska Hästavelsförbundet, 2024).

Norwegian Coldblooded Trotter is a breed that shares common origins with Norwegian Dola Horse but is specialized in trotting. In Sweden, a similar selection of North Swedish Horses used for harness racing led to the establishment of Swedish Coldblooded Trotters in 1964. Today, Norwegian and Swedish Coldblooded Trotters (Kallblodstravare) have a joint breeding programme (Det Norske Travselskap and Svensk Travsport, 2019). In 2023, there were 12,700 Norwegian Coldblooded Trotters, of which 5,800 were breeding animals, with a total of 555 coverings (Norsk Hestesenter, 2023). The same year, there were about 10,000 Swedish Coldblooded Trotters (Berglund *et al*, 2024), and 510 Swedish mares were covered (Svensk Travsport, 2024).

Gotland Pony (Gotlandsruss) is an old Swedish breed from the island of Gotland. In the 19th century, legal changes and land reforms led to a sharp decline in the Gotland pony population. Efforts for preservation began in the 1880s, and in the mid-20th century, a few stallions from foreign breeds were introduced to combat inbreeding (Andersson, 2016). In 2020, there were 5,700 Gotland Ponies (Hästnäringens Nationella Stiftelse, 2021), and 389 mares were covered in 2024 (Svenska Hästavelsförbundet, 2024).

Throughout the 20th century, the Norwegian Fjord Horse (Norsk Fjordhest) was selectively bred for its dun colour (White *et al*, 2024), which is now a distinguishing feature of the breed. Norwegian Fjord Horses were mixed with Dola Horses. Later, action was taken to remove Dola introgression from the breed (Olsen *et al*, 2020). Today, the Norwegian Fjord Horse has a breeding population of 2,200 in Norway (Norsk Hestesenter, 2023), and a decreasing number of coverings (280 in 2023), but additional related populations exist abroad (Olsen *et al*, 2020).

Organized breeding of Nordland/Lyngen Horses (Nordlandshest/Lyngshest) started around 1930, and the breed was primarily found in the northern part of Norway (White *et al*, 2024). The Nordland/Lyngen Horse was recognized as a breed in 1968 (White *et al*, 2024). One Finnhorse stallion was used in efforts to save the breed in 1979 (White *et al*, 2024). Today, there are approximately 2,900 horses of this breed in Norway, with a breeding population of 1,300 individuals (Norsk Hestesenter, 2023). The coverings have steadily decreased, reaching 175 in 2023 (Norsk Hestesenter, 2023).

Faroese Horses (Føroyskt Ross) were brought to the Faroe Islands by Norse settlers in the 9th and 10th centuries, and in the 19th century, were sold to the British Isles as pit ponies. Mass exportation and diminishing demand led to their almost extinction (Kjetså *et al*, 2024). When efforts to save the breed started in the 1960s, one stallion and four mares founded the current population (Kettunen *et al*, 2022). Today, the breed is critically at risk of extinction, with about 80 individuals, all living on the Faroe Islands (Kjetså *et al*, 2024).

There are some previous molecular population genetics studies of native Swedish, Norwegian and Faroese Horses (Bjørnstad *et al*, 2000; Bjørnstad and Røed, 2001; Bjørnstad and Røed 2002; Bjørnstad *et al*, 2003; Petersen *et al*, 2013; Andersson, 2016; Fegraeus *et al*, 2018; Sild *et al*, 2019; Velie *et al*, 2019b; Olsen *et al*, 2020; Kettunen *et al*, 2022; Smogeli, 2023; Joensen, 2024). However, these studies typically only targeted a few breeds, and several were done over a decade ago. Previous kinship assessments within these breeds primarily relied on pedigree records or limited numbers of DNA markers. Today, whole-genome sequencing (WGS) facilitates high-resolution examination of allelic variations and population structure. To the authors' knowledge, no previous studies examined the genetic diversity within and between the native Nordic horse breeds using whole-genome sequences.

The primary objective of this study was to conduct an in-depth analysis of genetic variation and kinship within and among the Nordic horse breeds using WGS data. A secondary objective was to compare genetic variation and diversity when using different reference genomes: the EquCab3.0 reference genome (NCBI, 2018) and the Nordic EquCab_Finn (Pokharel *et al*, 2024). This study encompasses measures such as fixation index, principal component analysis, effective population size, runs of homozygosity, heterozygosity, identity by state, and identity by descent. The results can be utilized when developing breeding strategies to maintain sustainable populations of native Nordic horse breeds.

Materials and methods

Sample collection and selection

This study involved whole blood samples from 190 individuals of the eight breeds: Swedish Ardennes, Dola Horse, North Swedish Horse, Coldblooded Trotter, Gotland Pony, Fjord Horse, Nordland/Lyngen Horses and Faroese Horse (Table 1). For most breeds, horses were sampled at breeding shows and similar gatherings of breeding horses in the summer and early fall of 2022. Some Coldblooded Trotter samples were collected in their stables or at a veterinary clinic. Most Nordland/Lyngen Horse samples were provided by Biobank AS in Hamar, Norway. Faroese Horse samples were provided by the breed association Felagið Føroysk Ross.

As far as possible, we aimed to include horses that had contributed or could contribute as breeding animals. Furthermore, we tried to include horses representing the variation in each population by including individuals with as low pedigree-based relatedness to each other as possible, which varied depending on breed. However, with the owners' consent, we were restricted to horses that were available for sampling. Both sexes were included for Dola Horses, North Swedish Horses, Coldblooded Trotters and Faroese Horses, whereas only stallions were available for the other breeds. A final selection of samples was based on the quality and concentration of extracted DNA.

DNA extraction, sequencing and data pre-processing

Genomic DNA (gDNA) was extracted from 190 whole blood samples using a QIASymphony instrument (QIAGEN, Hilden, Germany) and normalized to a concentration of 50ng/ml in low-salt Tris-EDTA buffer. Approximate DNA concentrations and OD260/280 ratios were measured using a Nanodrop instrument (Nanodrop Technologies, Wilmington, DE, USA). Further concentrations and quality control (QC) of gDNA was performed using an Agilent TapeStation (Agilent, Santa Clara, CA, USA). The SNP&SEQ Platform (National Genomics Infrastructure Sweden and Science for Life Laboratory, Uppsala, Sweden) prepared sequencing libraries by using TruSeq PCRfree DNA library preparation kit (Illumina Inc, San Diego, CA, USA). Paired-end sequencing was performed using v1.5 sequencing chemistry in two S4 flowcells on a NovaSeq 6000 sequencing instrument (Illumina Inc, San Diego, CA, USA).

The raw data was inspected with FastQC tools (Andrews, 2010) with standard options. Then, the GATK Best Practices workflow for short variant discovery in cohort analysis (Van der Auwera and O'Connor, 2020; Broad Institute, 2025) was used for data preprocessing. If no other information is given, the GATK v.4.0.8.0 default settings were used for all downstream tasks. The samples were first mapped on the Equine reference genome EquCab3.0 (NCBI, 2018) using bwa-mem (H. Li and Durbin, 2009). All reads were then assigned in a file to a single new read group with the AddOrReplaceReadGroups utility (Picard v. 2.23.4, (Broad Institute, 2020)). The duplicate reads were located and tagged with the MarkDuplicates utility (Picard v. 2.23.4, (Broad Institute, 2020)). Then, base quality score recalibration was done with GATK tools. First, the recalibration model was constructed using BaseRecalibrator. Known Single Nucleotide Polymorphisms (SNPs) listed in the horse genome were uploaded from the European Nucleotide Achive (ENA) through the European Variation Archive (EVA) (n.d.). Then, the scores were adjusted using the ApplyBQSR tools. Genome coverage was calculated using the samtools coverage command with the '-A -w 32' settings (Danecek *et al*, 2021). The average read depth of samples was 18.6X. To obtain the haplotypes in the cohort of all 190 samples, the SNPs and indels via local re-assembly of haplotypes were first detected for every sample separately using HaplotypeCaller tools from GATK with the '-ERC GVCF' setting (Poplin *et al*, 2018). Then, GenomicsDBImport was used to merge GVCFs from multiple samples to the GenomicsDB workspace. The joint genotyping of 190 samples was finally done using GenotypeGVCFs from GATK. The same pipeline was used for mapping and preprocessing the raw data on the EquCab_Finn genome. The average read depth of samples mapped on EquCab_Finn was 13X.

Quality control

Two rounds of QC were done using PLINK 2.0 (Chang *et al*, 2015; Purcell and Chang, 2017), one for all analyses and one for strict filtering needed for ROH and HET analysis. The resulting datasets are in this article called 'Filtered data' and 'Strictly Filtered data'. The raw data consisted of 25,988,853 variants. The first filtering round excluded all variants with a Phred Quality score less than 40 ('--var-min-qual 40'), giving the call an error rate of 0.01%. Further, read depth was required to be at least 8 with the option '--extract-if-info QD >= 8'. A minor allele count of 2 ('--mac 2') was imposed, reducing chances of a rarely observed allele being an error. Only autosomal variants were retained ('--chr 1-31'), leaving the Filtered EquCab3.0 mapped data with 21,655,806 variants. Additional filters to only retain SNPs ('--snps-only'), removing small indels from the data and removing all variants with any missing data ('--geno 0'), left the Strictly Filtered EquCab3.0 mapped data with 12,713,867 SNPs.

The same data QC was performed for the EquCab_Finn, and the EquCab3.0 mapped data. However, EquCab_Finn

Table 1. Overview of the samples per breed included in the study. [a], 14 CBT samples were collected in Norway and 14 in Sweden, but 18 had Norwegian registration numbers. [b], four NLH samples were collected in 2022; the rest were collected earlier and retrieved from the AS Biobank in Hamar, Norway.

Breed	Breed abbrev.	No. of horses	% males	Birth-years	Sampling Country	Sampling year
Swedish Ardennes	ARD	22	100	2017–2019	Sweden	2022
Dola Horse	DOL	30	30	2003–2021	Norway	2022
North Swedish Horse	NSH	30	67	2000–2019	Sweden	2022
Coldblooded Trotter	CBT	28	39	1998–2020	Norway/ Sweden[a]	2022
Gotland Pony	GTP	15	100	2009–2019	Sweden	2022
Fjord Horse	FJH	25	100	2007–2019	Norway	2022
Nordland/Lyngen Horse	NLH	30	100	1991–2021	Norway	2022[b]
Faroese Horse	FRH	10	67	1995–2021	Faroese Islands	2022

data had 29,251,175 raw mapped variants, the first filtering round yielded 22,451,933 variants for Filtered data, and Strictly Filtered data had 12,103,319 SNPs.

Analysis

Fixation index (F_{ST}) values were estimated using the Hudson method (Hudson *et al*, 1992; Bhatia *et al*, 2013), which is a refinement of Wright's F_{ST} (Wright, 1922) and corrects for potential biases introduced by rare variants. It calculates genetic differentiation between populations as a ratio of averages. Hudson's approach calculates averages of allele frequencies across populations before computing the F_{ST}. The general formula is:

$$\mathrm{Fst} = \frac{(\hat{p}_1-\hat{p}_2)^2-\left(\frac{\hat{p}_1(1-\hat{p}_1)}{n_1}+\frac{\hat{p}_2(1-\hat{p}_2)}{n_2}\right)}{\hat{p}_1(1-\hat{p}_1)+\hat{p}_2(1-\hat{p}_2)}$$

where $\hat{p}_1$ and $\hat{p}_2$ are allele frequencies in population one and two, respectively, while n_1 and n_2 are sample sizes of populations one and two. Input data was the Filtered data, which was then further filtered for the SNPs considered, a maximum sample- and variant-missingness of 10% ('--mind', '--geno'), and pruned with PLINK 2.0 linkage disequilibrium pruning tool ('--indep-pairwise') with an r^2 of 0.2 and distance of 50kb with a shift of 1kb. The analysis employed PLINK 2.0 '--FST' option, Hudson method, and all samples were included with the '--nonfounder' option. F_{ST} was calculated per family (breed) with the '--family' option.

Admixture analysis was done to determine the ancestry of samples, using the same data as for the F_{ST} analysis, and the R-package 'LEA' (v.3.18.0) (Frichot and François, 2015; Gain and François, 2021). The number of clusters (K) that minimized a cross-entropy criterion was used.

Principal component analysis (PCA) was used to reduce data complexity and capture important patterns (Greenacre *et al*, 2022). The Filtered dataset was pruned with PLINK 2.0 and the same settings as for F_{ST}, with a maximum number of two alleles and a variant genotype missingness of maximum 10%. PLINK 2.0 was used to do PCA ('--pca') and included all samples with the '--nonfounder' option, with variant filtering of a maximum of 10% missing data per variant ('--geno') and utilizing allele frequencies calculated ('--read-freq') for better estimations.

Strictly Filtered data was used for runs of homozygosity (ROH) estimation in PLINK 1.9 (Purcell and Chang, 2005; Chang *et al*, 2015) with the '--homozyg' option. The minimum number of SNPs within a run was set to 50 ('--homozyg-snp'), minimum run length was 300kbp ('--homozyg-kb'), maximum inverse density in a run was 50bp/SNP ('--homozyg-density'), maximum internal gap (kb) between two SNPs ('--homozyg-gap') was 1,000. Maximum number of heterozygotes in a scanning window was 1 ('--homozyg-window-het'), and scanning window size was 50 ('--homozyg-window-snp'). F_{ROH} is the measure of inbreeding based on ROH segments, calculated as:

$$\mathrm{F_{ROH}} = \sum \frac{\mathrm{k_{length}(ROH_k)}}{\mathrm{L}}$$

where k is the number of ROH identified in each sample in kb, and L is the total length of all autosomes. The total size of data mapped to autosomes was 2.28 GB for EquCab3.0, and 2.18 GB for EquCab_Finn.

Heterozygosity (HET) describes the total number of heterozygote variants in the population, and was calculated in PLINK using the method by Zhdanova and Pudovkin (2008). The percent frequency of observed and expected heterozygosity was calculated as:

$$\mathrm{Freq\,(HET)} = \frac{\text{(Counts of HET variants)}}{\text{(Total Count of Variants)}} \mathrm{x\ 100}$$

where HET variants were either Observed or Expected counts, and observed heterozygosity proportion was extracted using the '--het cols=+het' option in PLINK 2.0.

Historical effective population size (N_e) was estimated per breed using SNeP 1.1 (Barbato *et al*, 2015a; Barbato *et al*, 2015b) with default options. SNeP implements the linkage disequilibrium (LD) approach developed by Corbin *et al* (2012), and calculates N_e as:

$$\mathrm{N_e} = \frac{1}{4\mathrm{f}(\mathrm{r}^2)\times(\mathrm{c})}$$

where $f(r^2)$ is the average LD between pairs of SNPs (r^2) over a certain genetic distance, and c is the genetic distance between SNP pairs (in centiMorgans). This method relies on a decline in LD over increasing distances between loci, assuming faster LD decay in larger populations, and provides N_e at different time points. Default options were used for all chromosomes, minimum distance between SNPs was 50,000bp, and maximum was 4,000,000bp. SNeP imposes MAF filtering at 0.05, and maximum 100,000 SNPs analyzed per chromosome.

Identity by state (IBS) measures the proportion of identical alleles between two individuals at each SNP (Purcell *et al*, 2007) using the formula:

$$\mathrm{IBS} = \frac{\text{Number of IBS alleles}}{\text{Total number of genotyped SNPs between individuals}}$$

IBS is categorized into three states: IBS_0 when both alleles differ between two individuals, IBS_1 when one allele is identical, and IBS_2 when both alleles are identical between the two individuals. PLINK calculates overall IBS between individuals based on genome-wide IBS_0, IBS_1 and IBS_2 counts. For analysis of IBS, the Filtered data was pruned with PLINK 2.0 (Chang *et al*, 2015; Purcell and Chang, 2017) using the LD pruning tool ('--indep-pairwise'), with r^2 of 0.2, 50kb window length with 1kb increment, maximum number of alleles of 2, and 10% maximum variant genotype missingness. PLINK 2.0 estimated IBS using identity-by-descent option ('--genome') and the additional 'nudge' option to adjust the final estimates.

Identity by descent (IBD) measures variants that are likely from a common ancestor, and for this, the same options as for IBS were used. The fraction of pairwise comparisons (FPC) is the number of comparisons with nonzero values over the total number of comparisons. With this measure all animals are compared, meaning that breeds with fewer samples are more likely to have smaller FPC, and that breeds which share genetic material may have an FPC larger than the fraction of their own population size.

Results

Fixation index

The Dola Horse and North Swedish Horse had the lowest F_{ST} (0.030) (Table 2) when mapped on EquCab3.0. Also, the Coldblooded Trotter and North Swedish Horse had a low

F_{ST} of 0.051. Dola Horse and Coldblooded Trotter showed a moderate distance of 0.072, as did Swedish Ardennes and North Swedish Horse (0.078). At the upper end range of F_{ST} values, the Faroese Horse had F_{ST} values above 0.167 compared to all other breeds. Similar patterns were seen in F_{ST} values of EquCab_Finn data, with values within +/- 0.03 from those of EquCab3.0.

Admixture analysis

The lowest cross-entropy was found when including eight clusters (K) for the eight breeds, but the improvement compared with seven clusters was small (Supplemental Figure 1). The most noticeable admixture was seen between Dola Horse and North Swedish Horse, and Coldblooded Trotter and Dola Horse (Figure 1, Supplemental Tables 1 and 2). The lowest admixture was found in the Faroese Horse, the Gotland Pony, and the Nordland/Lyngen Horse. Minimal differences were seen between the EquCab3.0 and the EquCab_Finn data.

Principal component analysis

Together, the first three PCA components captured a substantial portion (45–47%) of genetic variation within each dataset (EquCab3.0 and EquCab_Finn). The two PCA plots (Figure 2) displayed a similar general pattern of breed similarities, but results deviated in some respects between the reference genomes. For example, Fjord Horse and Gotland Pony appeared more similar in the analysis of EquCab3.0 compared to EquCab_Finn data. The homogeneity of the first three principal components differed between breeds. The North Swedish Horse, Dola Horse and Coldblooded Trotter overlapped in both datasets, separating mainly across the first component.

Table 2. Fixation index (F_{ST}) values between breeds, from EquCab3.0 above and from EquCab_Finn data below the diagonal. Abbreviations as in Table 1.

	ARD	DOL	NSH	CBT	GTP	FJH	NLH	FRH
ARD		0.112	0.078	0.086	0.109	0.088	0.105	0.172
DOL	0.127		0.030	0.072	0.145	0.113	0.128	0.206
NSH	0.090	0.033		0.051	0.111	0.080	0.098	0.171
CBT	0.099	0.082	0.058		0.118	0.087	0.104	0.179
GTP	0.125	0.164	0.128	0.134		0.117	0.134	0.201
FJH	0.102	0.129	0.093	0.100	0.133		0.105	0.167
NLH	0.122	0.150	0.115	0.122	0.153	0.121		0.190
FRH	0.198	0.236	0.198	0.204	0.229	0.192	0.216	

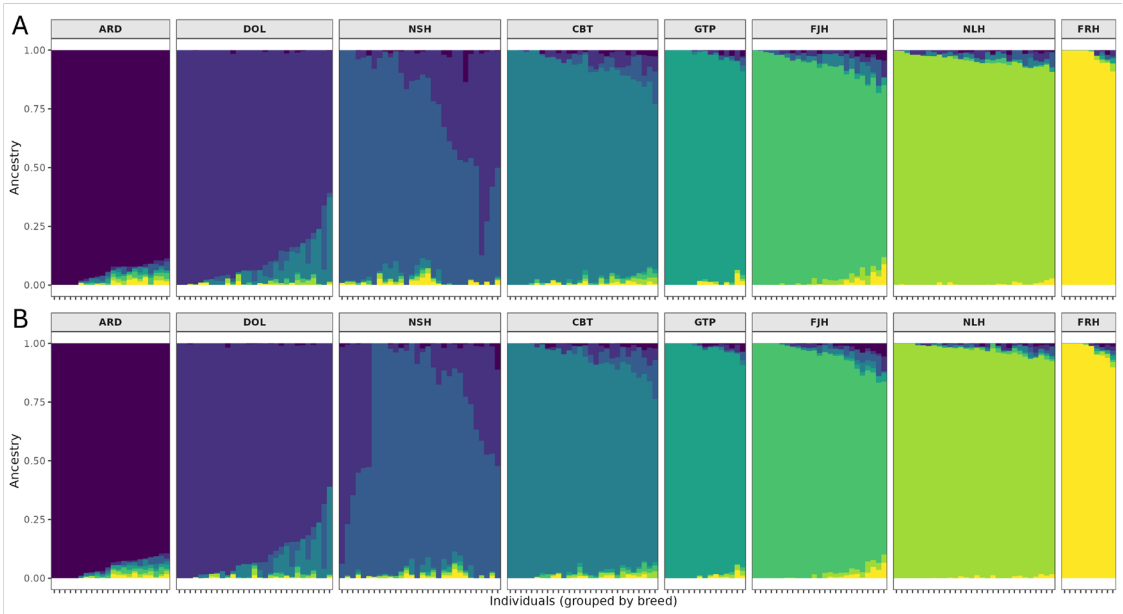


Figure 1. Admixture analysis assuming K = 8 clusters of A) EquCab3.0, and B) EquCab_Finn data. Abbreviations as in Table 1.

Runs of homozygosity

Violin plots (Figure 3) illustrate the distribution of inbreeding coefficients (F_{ROH}) for the breeds in EquCab3.0 and EquCab_Finn datasets. Faroese Horses consistently exhibited the highest inbreeding levels in both plots, with compact distributions (Table 3). Faroese Horse F_{ROH} ranged from 28.7% to 38.1% in EquCab3.0 data and from 27.2% to 36.6% in EquCab_Finn data. Swedish Ardennes showed the lowest inbreeding levels with the least variation between individuals (Figure 3). F_{ROH} for this breed ranged from 10.0% to 18.2% in In EquCab3.0 data, and from 8.9% to 16.8% in EquCab_Finn data.

The breeds Dola Horse, Coldblooded Trotter, Fjord Horse and North Swedish Horse showed broader distributions of F_{ROH}. The Dola Horse, for instance, ranged from 7.7% to 30.4% in EquCab3.0 data and from 6.4% to 27.8% in EquCab_Finn data. The individual with the lowest F_{ROH} of about 1% was a North Swedish Horse from a Dola Horse sire approved for breeding in the North Swedish Horse breed. Nordland/Lyngen Horse and Gotland Pony displayed relatively narrow distributions.

In the EquCab3.0 data, the Dola Horse, Gotland Pony, and Faroese Horse showed larger contributions of ROHs > 4Mb to the total F_{ROH} than the other breeds. In the EquCab_Finn data, the contributions from ROH > 4Mb were considerably lower for all breeds, however (Table 3).

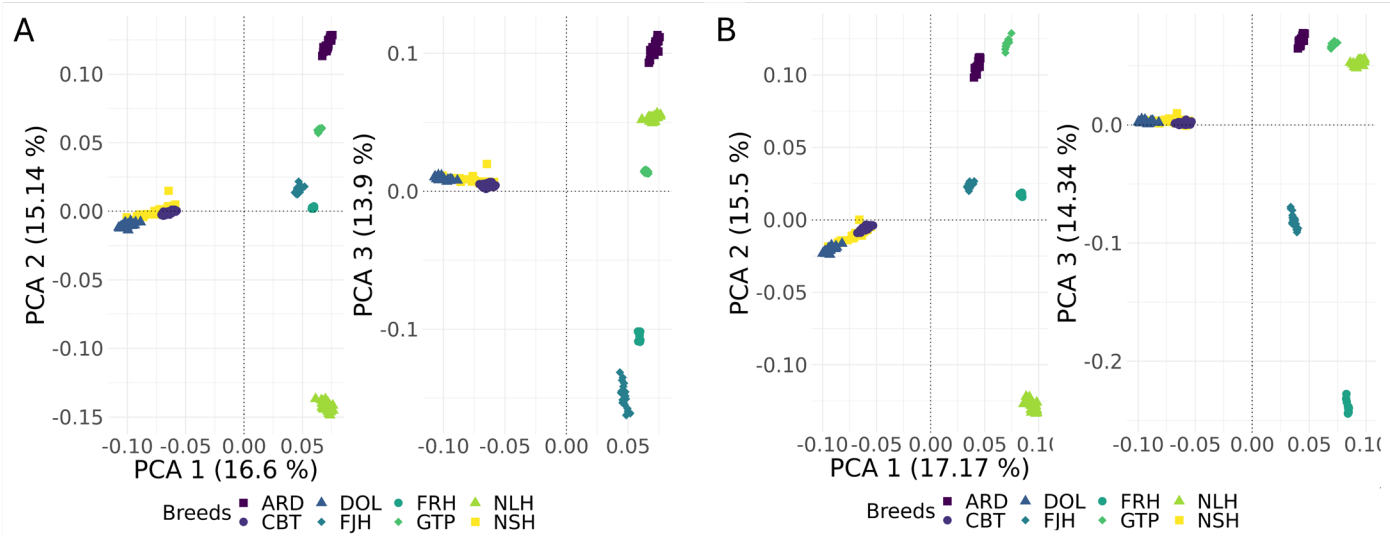


Figure 2. Principal component analysis (PCA) plot of (A) EquCab3.0, and (B) EquCab_Finn data, comparing PCA1 vs. PCA2 and PCA1 vs. PCA3. For EquCab3.0 data, PCA1 explained 16.6% variation, PCA2 15.1%, and PCA 3 13.9%. For EquCab_Finn data, PCA1 explained 17.7% variation, PCA2 15.5%, and PCA3 14.43%. Abbreviations as in Table 1.

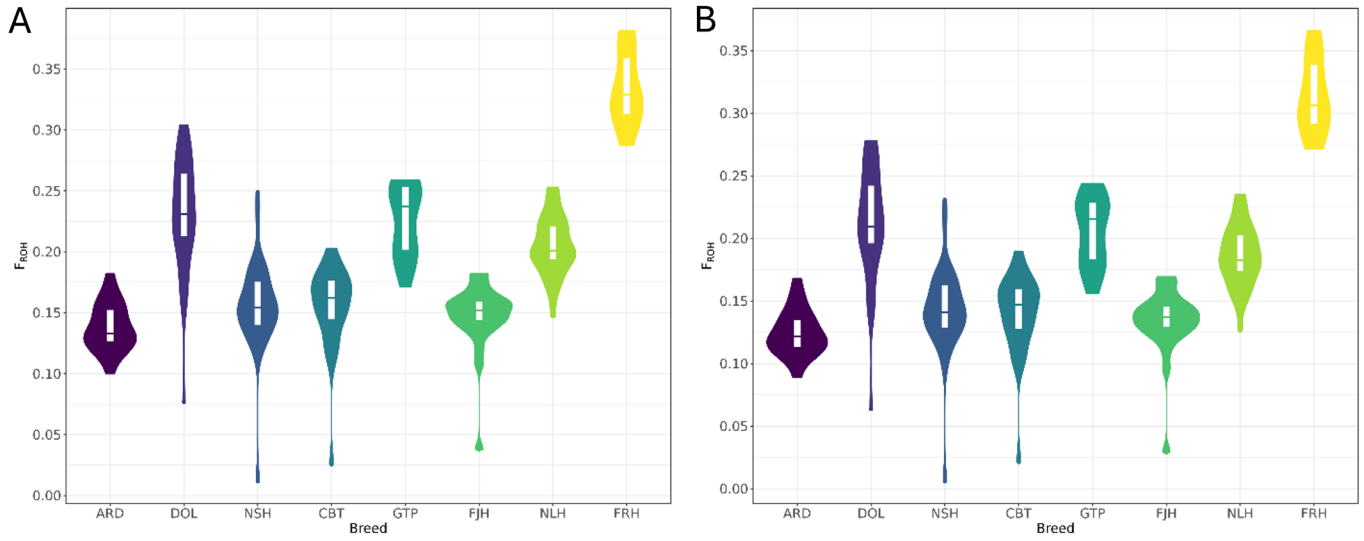


Figure 3. Distribution of inbreeding coefficients (F_{ROH}) for (A) EquCab3.0 and (B) EquCab_Finn data within breed. Abbreviations as in Table 1.

Table 3. Inbreeding coefficients (F_{ROH}) values for each breed in EquCab3.0 and EquCab_Finn data. Breed abbreviations as in Table 1. [a], The categories include the lower limit but exclude the upper limit.

		Mean F_{ROH} (%) per length category (Mb)[a]				Total F_{ROH} (%)		
	Breed	**0.3–1**	**1–2**	**2–4**	**> 4**	**Min**	**Max**	**Mean**
EquCab3.0	ARD	5.8	3.3	2.9	1.8	10.0	18.2	13.8
	DOL	5.8	4.7	6.3	6.2	7.7	30.4	23.0
	NSH	5.1	3.2	3.9	3.4	1.2	24.9	15.6
	CBT	5.2	3.4	3.6	3.3	2.6	20.3	15.5
	GTP	5.7	4.3	6.3	6.1	17.1	25.9	22.5
	FJH	5.4	2.9	3.2	3.2	3.8	18.2	14.7
	NLH	6.3	4.6	5.3	4.3	14.7	25.3	20.5
	FRH	6.7	5.4	8.3	13.0	28.7	38.1	33.4
EquCab_Finn	ARD	7.1	3.5	1.7	0.2	8.9	16.8	12.5
	DOL	9.3	6.5	4.3	1.0	6.4	27.8	21.2
	NSH	6.8	4.2	2.7	0.6	0.6	23.1	14.3
	CBT	7.0	4.0	2.5	0.5	2.1	19.0	14.0
	GTP	8.9	6.3	4.4	1.0	15.6	24.4	20.5
	FJH	6.8	3.8	2.2	0.5	2.9	17.0	13.3
	NLH	8.9	5.6	3.4	0.6	12.6	23.5	18.6
	FRH	11.5	9.7	7.4	2.7	27.2	36.6	31.4

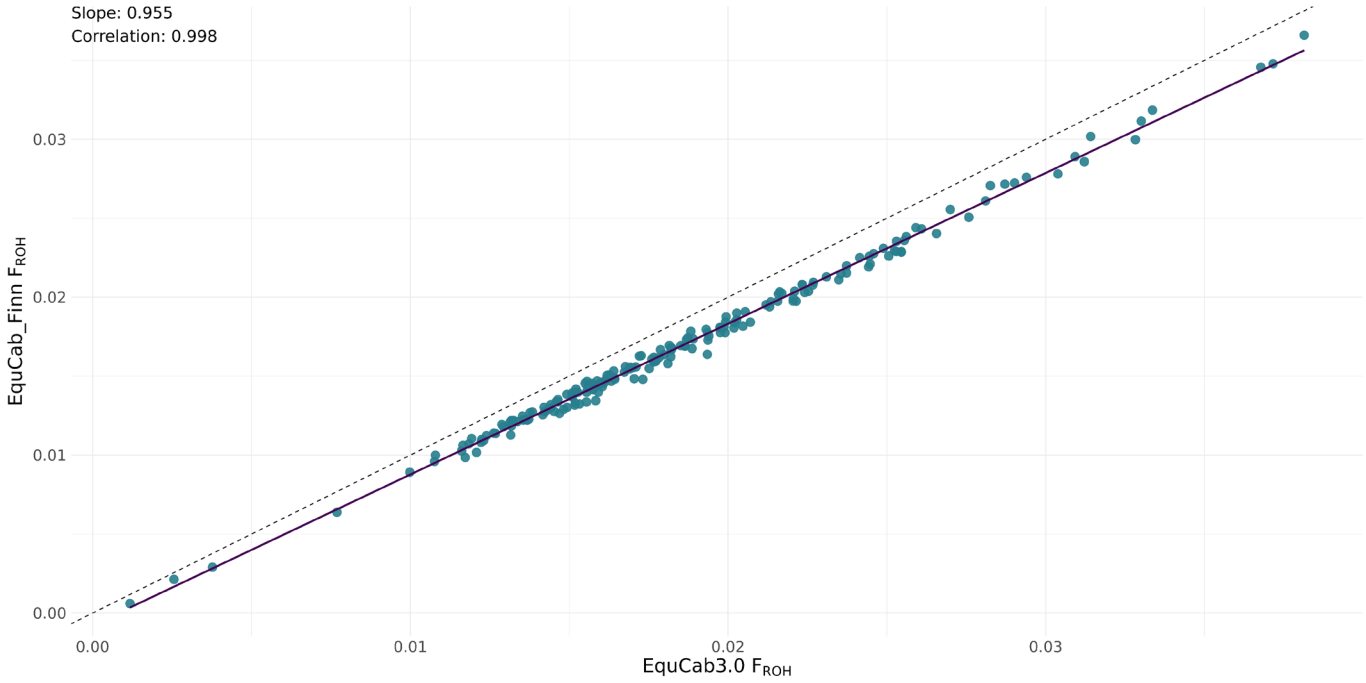


Figure 4. Comparison of inbreeding coefficients (F_{ROH}) values between EquCab3.0 (x-axis) and EquCab_Finn (y-axis) datasets.

The correlation was strong (0.998), as was the regression coefficient (0.955) between individual total F_{ROH} in the EquCab_Finn data and EquCab3.0 data, however, with Faroese Horses showing highest overall inbreeding levels and Swedish Ardennes consistently the lowest (Figure 4).

Heterozygosity

In both EquCab3.0 and EquCab_Finn data, Swedish Ardennes had the highest mean observed heterozygosity (Table 4) of close to 16%, and Faroese Horse had the lowest of less than 12%. Despite this, observed heterozygosity exceeded expected heterozygosity, which was the case for all the breeds. The Dola Horse demonstrated the second lowest mean observed heterozygosity of 13.4% in both datasets, whereas the Gotland Pony, Coldblooded Trotter, Fjord Horse, Nordland/Lyngen Horse and North Swedish Horse all showed estimates between 14.6% and 15.3%.

Table 4. Expected (Exp.), and observed heterozygosity (mean, maximum and minimum) in EquCab3.0 and EquCab_Finn data.

Breed	EquCab3.0				EquCab_Finn			
	Exp. (%)	Mean (%)	Max (%)	Min (%)	Exp. (%)	Mean (%)	Max (%)	Min (%)
Swedish Ardennes	15.6	16.2	17.1	14.5	14.9	16.0	16.8	14.4
Dola Horse	13.1	13.4	15.4	7.1	12.7	13.4	15.3	7.4
North Swedish Horse	14.8	14.9	16.5	12.0	14.3	14.8	16.4	12.0
Coldblooded Trotter	14.9	15.3	16.3	13.9	14.3	15.3	16.2	14.0
Gotland Pony	14.1	14.6	16.2	10.9	13.6	14.6	16.1	11.1
Fjord Horse	15.0	15.3	16.6	9.7	14.4	15.2	16.5	9.8
Nordland/ Lyngen Horse	14.3	14.7	15.8	13.5	13.8	14.7	15.7	13.6
Faroese Horse	10.9	11.7	13.0	10.1	10.6	11.8	13.0	10.3

Historical effective population size

N_e for the various horse breeds based on EquCab3.0 and EquCab_Finn mappings were similar, with slight differences between the two datasets. The estimated N_e generally increased further back in time (not shown). North Swedish Horse had the highest N_e of 96 in EquCab3.0 data and of 99 in EquCab_Finn when looking at the effective population size estimated over the past 13 generations (Table 5). The lowest corresponding N_e was found for Faroese Horse, at 23 and 24 for EquCab3.0 and EquCab_Finn data, respectively. EquCab_Finn data gave slightly higher estimates than EquCab3.0 data for all studied breeds.

Table 5. Effective population size estimated over the past 13 generations per breed in EquCab3.0 and EquCab_Finn data.

Breed	EquCab3.0	EquCab_Finn
Swedish Ardennes	85	87
Dola Horse	69	72
North Swedish Horse	96	99
Coldblooded Trotter	74	76
Gotland Pony	44	46
Fjord Horse	75	78
Nordland/ Lyngen Horse	67	70
Faroese Horse	23	24

Identity by state

Differences in IBS values between genome assemblies were small and consistent, with EquCab_Finn data presenting from 0.3 to a few percentage points higher estimates than EquCab3.0 data (Table 6). The highest mean IBS was found in Dola Horses, at 88% in EquCab3.0 data and 89% in EquCab_Finn data. However, the highest IBS of more than 94% was found for a Faroese Horse. The lowest mean IBS was found in Gotland Pony and Swedish Ardennes at 86% in EquCab3.0 and 88% in EquCab_Finn datasets.

Identity by descent

EquCab3.0 and EquCab_Finn data heatmaps show similar overall patterns of IBD between groups of individuals belonging to different breeds (Figure 5). However, proportions of IBD showed slightly different clustering within breeds and between specific individual pairs in the EquCab3.0 data compared to EquCab_Finn data.

Mean IBD for Swedish Ardennes was 17% in EquCab3.0 data, and slightly lower (15%) in the EquCab_Finn data, whereas it was 32% for Dola Horse in both reference genomes (Table 7). For Faroese Horse, mean IBD remained at 40% across both genomes. The largest difference between the reference genomes in mean IBD (26% vs 21%) was seen for Fjord Horse. For all breeds, pairwise comparisons gave lower FPC in EquCab3.0 data than in EquCab_Finn data (Table 7). Differences in IBD between EquCab3.0 and EquCab_Finn were small relative to the standard deviation of IBD.

Discussion

Historical background and use of the native Nordic horse breeds differed, but they all played an important role in developing the Nordic countries. With modernization of transport, agriculture and forestry, the roles of horses changed. Facing decreased population sizes, it is urgent to support sustainable breeding strategies for the breeds, and the first step is to characterize standing genetic variation. This study is the first to analyze and compare WGS data for all native Norwegian and Swedish horse breeds. In addition, we included the critically endangered Faroese Horse. Some other Nordic breeds, mainly Icelandic Horse and Finnhorse, have previously been more studied (Kierkegaard *et al*, 2020). Native Nordic breeds are not closely related to Thoroughbreds (Petersen *et al*, 2013). Therefore, the novel comparison between using the EquCab3.0 (Thoroughbred) reference genome and the newly developed EquCab_Finn genome based on the Nordic breed Finnhorse was of special interest.

Table 6. Identity by state values for the EquCab3.0 and the EquCab_Finn datasets.

	Breed	Mean (%)	Median (%)	Max (%)	Min (%)
EquCab3.0	Swedish Ardennes	85.7	85.6	90.0	84.7
	Dola Horse	88.1	87.5	92.1	85.5
	North Swedish Horse	87.0	86.9	93.0	85.2
	Coldblooded Trotter	87.8	88.0	92.2	85.2
	Gotland Pony	85.6	85.5	90.1	84.8
	Fjord Horse	86.1	85.9	92.4	85.3
	Nordland/ Lyngen Horse	86.5	86.1	92.5	85.3
	Faroese Horse	86.1	85.5	94.2	85.2
EquCab_Finn	Swedish Ardennes	87.5	87.5	91.3	86.7
	Dola Horse	89.8	89.3	93.2	87.4
	North Swedish Horse	88.8	88.7	94.0	87.2
	Coldblooded Trotter	89.4	89.6	93.3	87.1
	Gotland Pony	87.5	87.4	91.4	86.8
	Fjord Horse	87.9	87.8	93.5	87.0
	Nordland/ Lyngen Horse	88.3	88.0	93.5	87.2
	Faroese Horse	88.0	87.5	94.9	87.1

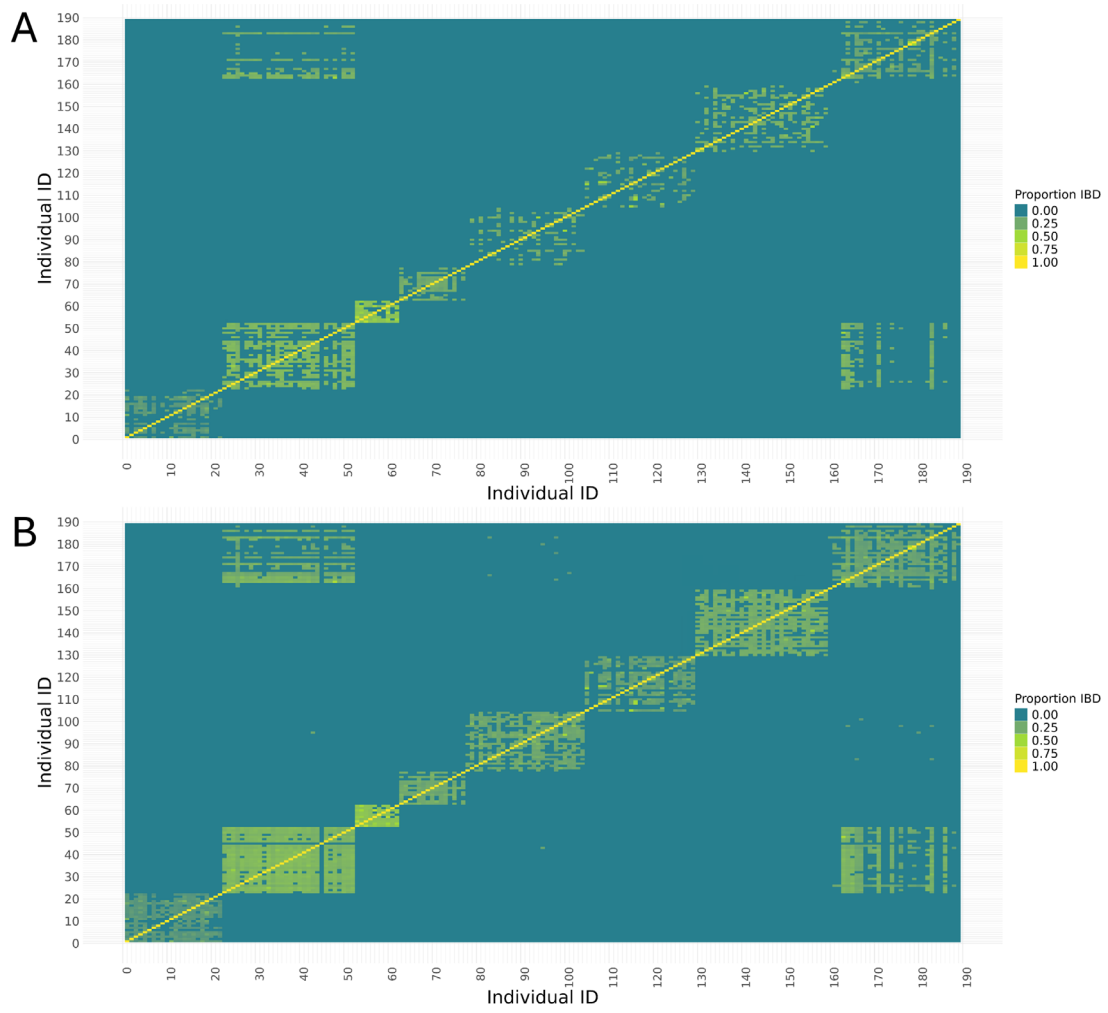


Figure 5. Heatmap of IBD relationships between individuals for (A) EquCab3.0, and (B) EquCab_Finn data, with each individual's relationship with itself on the diagonal. Stronger yellow indicates higher IBD. Individual IDs 0-22 are Swedish Ardennes, 23-52 Dola Horses, 53-62 Faroese Horses, 63-77 Gotland Ponies, 78-104 Coldblooded Trotters, 105-129 Fjord Horses, 130-159 Nordland/Lyngen Horses, and 160-189 North Swedish Horses.

Table 7. Mean identity by descent (IBD) in percentage based on non-zero IBD pairs with standard deviations (SD) in brackets, and fraction of pairwise comparisons with nonzero IBD (FPC), in percentage.

	EquCab3.0		EquCab_Finn	
Breed	**Mean (SD)**	**FPC**	**Mean (SD)**	**FPC**
Swedish Ardennes	17 (3.5)	2	15 (3.3)	4
Dola Horse	32 (2.2)	19	32 (2.6)	31
North Swedish Horse	26 (4.4)	7	25 (4.5)	15
Coldblooded Trotter	27 (3.9)	11	24 (3.4)	40
Gotland Pony	26 (2.1)	2	25 (2.4)	3
Fjord Horse	26 (7.1)	2	21 (5.9)	5
Nordland/ Lyngen Horse	28 (3.2)	5	27 (2.6)	11
Faroese Horse	40 (6.4)	7	40 (6.2)	9

EquCab3.0 vs. EquCab_Finn

Comparison of both admixture results, F_{ST} values, and PCA analysis from the different reference genomes showed that EquCab_Finn genome in general aligned closely with EquCab3.0 in reflecting the genetic structure across the breeds, although subtle variations between datasets were seen. Slight differences in annotation and down-stream analyses of population genetics depending on reference genome were previously reported, e.g. cattle, canines and fish (Gopalakrishnan *et al*, 2017; Weldenegodguad *et al*, 2019; Lloret-Villas *et al*, 2021; Thorburn *et al*, 2023).

Correlations between F_{ROH} estimates from using the two reference genomes in our study indicated a nearly perfect linear relationship. However, the regression coefficient pointed to slightly higher F_{ROH} from EquCab3.0 than from EquCab_Finn. The Finnhorse assembly likely called alleles specific to the Nordic breeds, detecting more variation and breaking up some of the EquCab3.0 derived ROHs. This was supported by the considerably smaller contribution of long (> 4Mb) ROH to the F_{ROH} in the EquCab_Finn than in the EquCab3.0 data.

Differences between the reference genomes in observed heterozygosity were very small, whereas expected heterozygosity was somewhat lower in the EquCab_Finn data. This may be affected by the relatively small number of samples per breed, or by additionally mapped variants in EquCab_Finn having more extreme allele frequencies, which would be likely for regions specific to Nordic breeds. Similar but more intense tendencies were previously seen in canines, for reference genomes of the wolf and the dog (Gopalakrishnan *et al*, 2017).

The N_e estimates indicate an approximate number of effective ancestors that contributed to the breed of today. Although the values should be interpreted with care (Adepoju *et al*, 2024), they can give an indication of relative differences between the breeds. One should keep in mind that formal breed formation and start of more regulated breeding for most breeds did not take place until the late 19th or early 20th century, and that generation intervals are typically around ten years or more in horse breeds (Viklund *et al*, 2011; Olsen *et al*, 2020). The EquCab_Finn data showed a slightly higher estimated historical N_e in all the breeds, likely due to more heterozygous variants, as also seen in the test of heterozygosity (Table 4).

The mean IBDs were, in most cases, similar between the two assembly genomes, while the proportion of animals having non-zero IBD values with other individuals increased for several breeds when using the EquCab_Finn assembly compared to EquCab3.0.

Overall, we found minor to moderate differences when using the two reference genomes. Pokharel *et al* (2024) reported that EquCab_Finn shared about 95% of the genomic features of EquCab3.0. The type of analysis and comparisons between breeds in our study are likely less sensitive to this difference than, for example, detecting unique, rare genetic variants.

Diversity within Norwegian, Swedish and Faroese breeds

Swedish Ardennes counted high numbers during the first half of the 20th century (Carlström *et al*, 1946), before a drastic reduction when tractors replaced horses in agriculture. The breed showed the lowest F_{ROH} and IBD values among the breeds, possibly remnants of the previously large population. However, heterozygosity at 16% suggests low genetic diversity. The breed's estimated N_e of 85–87, 13 generations back, was lower than the estimate of 227 in 2002 based on rather low-depth pedigree data (Siekas, 2006).

The dispersion of Dola Horse FROH around the mean of 21–23% indicated individual differences in genetic variability. The heterozygosity of 13% reflects high genetic diversity loss, and the mean IBD of 32% indicates few common ancestors for the population. The N_e 13 generations ago of 69–72 was between previously pedigree-based estimates of 51–151 for the years 2010–2015 (Melheim, 2017), and 152 for the years 1990–1999 (Olsen *et al*, 2010). Melheim (2017) estimated pedigree inbreeding coefficients of 8–13% for the Dola Horse.

The North Swedish Horse had a lower F_{ROH}, dispersed around 14–16%, like for Coldblooded Trotters. The ongoing use of approved Dola Horses in breeding of the North Swedish Horse likely contributes to an increased diversity, and to the highest estimate (96–99) of N_e 13 generations ago in this study. North Swedish Horses were sampled in Sweden under the Swedish breed name, but in hindsight, at least one individual could also have been defined as a Dola Horse.

The estimated N_e 13 generations ago for the Coldblooded Trotter of 74–76 was closer to that of the Dola Horse than

the North Swedish Horse. Declining N_e has been estimated based on pedigree data for the Coldblooded Trotter, with an estimate of 40–50 for the years 2012–2015 (Leroy *et al*, 2020). This breed, which is under intense selection for trotting performance, showed low genetic diversity in several measures, including a high IBD of 24–27% and low heterozygosity (15%). A mean pedigree-based inbreeding coefficient of 8.3% was estimated for Coldblooded Trotters born in 2021 by Berglund *et al* (2024). Velie *et al* (2019b) estimated mean F_{ROH} based on SNP data of 9.6% in Norwegian, and 8.7% in Swedish Coldblooded Trotters, born 2000–2009.

A severe genetic bottleneck in the late 19th and early 20th centuries left its mark on the Gotland Pony breed, despite different measures taken since then to improve the situation (Svenska Russavelsföreningen, 2019). Their F_{ROH} clustered around 21–23%, with little individual variation. Also, their mean observed heterozygosity of 15% and IBD value of 25–26% suggest low genetic diversity. The estimated N_e 13 generations ago was as expected low (44–46). A pedigree-based mean inbreeding coefficient of 11% and N_e of 67 was previously estimated for Gotland Ponies by Andersson (2016).

Substantial individual variation in FROH was seen for the Fjord Horse, for which there is some exchange with foreign Fjord Horse populations. Also, for this breed, the heterozygosity (15%) indicated low genetic diversity, The estimated N_e size 13 generations ago of 75–78 concurs with previous estimates of 63 on molecular data from 2015 and 71 on pedigree data (Olsen *et al*, 2020). The latter estimated a mean pedigree-based inbreeding of 7.7%.

Nordland/Lyngen Horses' F_{ROH} clustered around 19–21%, and their heterozygosity (15%), and IBD (27–28%) indicated low genetic variation. The N_e 13 generations ago of 67–70 was within the range of previous pedigree-based estimates of 20–75 in 1988–1991, and 121–176 in 2012–2015 (Leroy *et al*, 2020).

The high F_{ROH}, clustering around 31–33%, in Faroese Horse concords with the pedigree inbreeding coefficient of 27% estimated by Kettunen *et al* (2022) for Faroese Horses born in 2016. Faroese Horses displayed the highest IBD value (40%), lowest observed heterozygosity (12%), and smallest historical N_e among the breeds. This agrees with the breeding history, including a relatively recent extreme bottleneck and a small, isolated population (Kettunen *et al*, 2022).

Population subdivision has been shown to influence estimation of genetic diversity through a deficit in heterozygotes relative to Hardy-Weinberg expectations when animals are sampled from different inbred lines (Wahlund effect) (De Meeûs, 2018). In larger breeds under divergent selection for different breeding goals, for example, American Quarter Horses (Petersen *et al*, 2014), genomic analyses showed substructures within the breed. Population substructures due to favoured stallions and admixture with other breeds were shown for Franches-Montagnes horses (Gmel *et al*, 2024). In the Costa Rican Paso horse, which started as a synthetic breed and still shows subdivision, reduction of heterozygotes, likely due to the Wahlund effect, was reported in a study by Domínguez-Viveros *et al* (2024). However, we found no indications of clear substructures in the numerically small native breeds in the present study based on expected vs. observed heterozygosity, PCA, IBD, or admixture analysis. This agrees with their known breed origins and present breeding programmes. Notably, even though all breeds in this study showed loss of genetic diversity, they exhibited somewhat higher observed than expected heterozygosity, which reflects the breed management, aiming to avoid close inbreeding.

Comparisons with results from other breeds

The number of studies of horses using WGS data is increasing, but still limited. Several studies based on SNP marker data are available, however. While F_{ST} and PCA values of other breeds and species do not convey much relevant information for comparison, the contrary is true for heterozygosity and F_{ROH} measures, although different settings may influence comparisons across studies. We did not filter for minor allele frequency or LD in the data used to detect ROHs, following recommendations by Meyermans *et al* (2020). Selection of individuals to be included may also influence the estimates, and horses in the present study were selected with the intent to capture as much of the present diversity as possible.

Compared with other breed estimates, the Nordic breeds in our study had low heterozygosity (12–16%), in some cases due to past bottlenecks and in other likely strong selection. For example, Y. Li *et al* (2022) reported observed heterozygosity of 18–32% in WGS data on Chinese Indigenous horse breeds. In analysis of SNP data, observed heterozygosity was 33% in Icelandic Horse and 34% in Exmoor Pony, two native northern European breeds (Sigurðardóttir *et al*, 2024) that have experienced fluctuating population sizes. Still, estimates of F_{ROH} between 8–20% for Icelandic horse and 12–27% for Exmoor Pony (Sigurðardóttir *et al*, 2024) means that those breeds would be placed in the lower to mid-range of results for breeds in the present study.

As a comparison, F_{ROH} was estimated to be 26% in a Thoroughbred population using WGS data (Chen *et al*, 2023), which was higher than for the Swedish and Norwegian breeds in our study. A large international population, but also few founder stallions, a strong male selection, and so-called line breeding characterize the Thoroughbred (McGivney *et al*, 2020). Corbin *et al* (2010) estimated the historical population size of Thoroughbreds 20 generations ago to be 100, with an increase in recent times based on SNP data. This is close to the highest estimate for the North Swedish Horse, found in the present study.

Genetic similarity between the studied breeds

The structure of the genetic similarity between horse breeds based on F_{ST} and PCA in the present study was at least partly expected, based on known breed history. Swedish Ardennes has its roots in Belgium, and F_{ST} for Swedish Ardennes showed moderate differentiation from other breeds, with slightly lower differentiation from North Swedish Horse (0.08), possibly a remnant of past crossbreeding. The admixture analysis showed individuals with ancestry from both the North Swedish and Dola breeds, and to a lesser extent from the Dola and Coldblooded trotter. This was expected based on pedigree information and that some crossbreeding is allowed, as previously described. The Dola Horse showed less differentiation based on F_{ST} from Coldblooded Trotter (0.07) and North Swedish Horse (0.03) than from the other breeds, and clustered closely with North Swedish Horse in the PCA,

as expected due to both recent and historical exchange of genetic material between these breeds (Norsk Hestesenter, 2023). A slightly larger distance from the Coldblooded Trotter compared with the North Swedish Horse, despite their common origin, may be due to loss of genetic variation. This phenomenon of breeds with a low genetic variation that seems far distant from other breeds is seen, for example, in Friesian Horses (Schurink *et al*, 2019).

North Swedish Horse displayed moderate F_{ST} values, with lower differentiation from the working horse breeds Dola Horse (0.03), Swedish Ardennes (0.08) and Norwegian Fjord Horse (0.08), and the Coldblooded Trotter (0.05). This confirms previous F_{ST} estimated between North Swedish horse and Coldblooded Trotters at 0.04 (Fegraeus *et al*, 2018), and 0.07 (Velie *et al*, 2019a). Fegraeus *et al* (2018) also estimated a similar F_{ST} value (0.08) between Coldblooded Trotter and Dola Horse. Velie *et al* (2019a) revealed evidence of introgression of genetic variants from Standardbreds into Coldblooded Trotters, before parentage verification was introduced in the 1960s (Sweden) and 1970s (Norway) (Det Norske Travselskap and Svensk Travsport, 2019).

Gotland pony exhibited moderate F_{ST} values to other breeds, and the PCA plot showed a well-separated cluster for Gotland Ponies, signifying its genetic distinctiveness. None of the foreign breeds, such as Welsh Ponies introduced to increase the genetic diversity in Gotland Ponies (Andersson, 2016), were included in the present study and thus, it was not possible to estimate their relatedness.

Previously reported F_{ST} for Fjord Horse to Coldblooded Trotter (0.09) and to Dola Horse (0.15) (Fegraeus *et al*, 2018) were similar to those in our study (0.09 and 0.11, respectively). The PCA analysis for the Fjord Horse showed some dispersion in PCA3, indicating subtle genetic variation within the population. Only Norwegian Fjord Horses were included, and the results may have been somewhat altered if Fjord Horses had been sampled abroad.

The Nordland/Lyngen Horse showed moderate F_{ST} differentiation from all other breeds, similar to those by Fegraeus *et al* (2018) between Nordland/Lyngen Horse and Fjord Horse (0.13), Coldblooded Trotters (0.14), and Dola Horse (0.22). The high genetic differentiation of the Faroese Horse from all other breeds, with the lowest F_{ST} of 0.17 to the Fjord Horse, can be ascribed to the breed's history with a severe bottleneck and isolation, and, thus, genetic drift.

For future studies, adding breeds such as Thoroughbred, Standardbred, or Arabian horses would make the results more relatable to other breeds. It would also be of interest to include breeds from other Nordic countries, subpopulations of Swedish and Norwegian breeds abroad, and draught horses of a similar origin to the Swedish Ardennes. Studying genetic similarity between the Nordic, and the British and Baltic breeds would be of interest as there are historical connections with these regions. In addition, analysis of structural variations in the studied breeds may reveal additional aspects of their relatedness within and across populations. Results for the two reference genomes support that an equine pan-genome, which is being developed (Stroupe *et al*, 2024), will be a strong advancement for comparing genetic diversity on a detailed level across breeds with different origins.

Limitations

There is a lack of golden standards regarding many of the analyses included in this study, creating some difficulties in comparing results with previous studies. Further, estimation of effective population size based on linkage disequilibrium is challenging, as the population history and structure can impact the results (Ryman *et al*, 2019; Adepoju *et al*, 2024). Recently, Manunza *et al* (2025) suggested a minimum of 50 individuals to determine N_e using LD-based methods. However, they studied moderate marker density array data and breeds with much larger population sizes than in our study. Despite few individuals being sampled per breed, we assume we could capture a large proportion of genetic variations because the populations are so small, and several have undergone relatively recent genetic bottlenecks. This is supported by the estimation of the necessary sample size for F_{ST} analysis, when using high marker density, presented by Nazareno *et al* (2017).

Conclusions

While the patterns of genetic similarity across breeds appeared consistent, the choice of reference genome (EquCab_Finn or EquCab3.0) affected the number of samples showing genetic relationships and overall IBD values. The minor to moderate differences between results from using the two reference genomes suggest that using the EquCab_Finn genome improves the precision of genomic mapping and, thus, understanding of the genetic diversity and relationships between and within Nordic horse breeds. Further, different analyses consistently showed reduced genetic diversity in all breeds studied. The current population estimates tell us their present status. However, their future also depends on the possibility of maintaining or increasing population sizes, which may be challenging for some breeds. Maintaining genetic diversity in these breeds is essential for the survival of the breeds and thereby for preserving a significant part of Nordic heritage for future generations.

Supplemental data

Supplemental Figure 1. Cross-entropy from admixture analysis of the eight breeds across K values of 2-12 for A) EquCab3.0, and B) EquCab_Finn mapped data

Supplemental Table 1. Admixture contribution of the different breeds to each breed in the EquCab3.0 data

Supplemental Table 2. Admixture contribution of the different breeds to each breed in the EquCab_Finn data

Acknowledgements

This project was funded by the Swedish-Norwegian Foundation for Equine Research (H-21-47-653). The authors thank the breed associations and horse owners for allowing sample collection and contributing veterinarians for collecting samples. We acknowledge the AS Biobank in Hamar, Norway, and the Faroese Horse Association for providing samples, and the Animal Genetics Laboratory at the Swedish University of Agricultural Sciences, for DNA extraction. We thank the Swedish Trotting Association, Svenska Hästavelsförbundet, and the Norwegian Equine Association for providing pedigree

information. The sequencing was performed by the SNP&SEQ Technology Platform in Uppsala, which is part of the National Genomics Infrastructure (NGI) Sweden and Science for Life Laboratory and also supported by the Swedish Research Council and the Knut and Alice Wallenberg Foundation. The computations and data handling were enabled by resources provided by the National Academic Infrastructure for Supercomputing in Sweden (NAISS), partially funded by the Swedish Research Council through grant agreement no. 2022-06725, and by the Orion High-Performance Computing Center (OHPCC) at the Norwegian University of Life Sciences.

Author contributions

NAS – analysis, interpretation of results, writing (original draft), visualization; IS – data pre-processing, bioinformatics methodology; SKR – sample contribution, sample selection; MK – project initiation. JK and KP – access to the EquCab_Finn reference genome. TS – sample collection; SM – project initiation, sample selection, methodology, interpretation of results; SE – project initiation, main applicant for Swedish funding, sample collection, sample selection, methodology, interpretation of results; PB – project initiation, main applicant for Norwegian funding, methodology, sample selection, supervision, interpretation of results. All authors reviewed, edited, and approved the final manuscript.

Conflict of interest statement

The authors declare no conflict of interest.

Ethics statement

The study was reviewed and approved by the Swedish regional ethics committee on animal experiments Uppsala djurförsöksetiska nämnd (ethical permit Dnr 5.8.18-05055/2019) and the Norwegian Food Safety Authority Mattilsynet (ethical permit FOTS ID 29635). Written informed consent was obtained from the owners for the participation of their animals in this study.

Availability of data

The data for this study are deposited in the European Nucleotide Archive (ENA) at EMBL-EBI under accession number PRJEB108320 (https://www.ebi.ac.uk/ena/browser/view/PRJEB108320). The code scripts for the analysis are published on the GitHub page: https://github.com/NattyandMinnie.

References

Adepoju, D., Ohlsson, J. I., Klingström, T., Rius-Vilarrasa, E., Johansson, A. M., and Johnsson, M. (2024). Population history of Swedish cattle breeds: estimates and model checking. *bioRxiv*, 1-28. doi: https://doi.org/10.1101/2024.10.03.616479

Andersson, L. (2016). Analysis of inbreeding in the Swedish Gotland pony using pedigree information and microsatellite markers. M.Sc. Thesis. Swedish University of Agricultural Sciences, Uppsala.

Andrews, S. (2010). FastQC: a quality control tool for high throughput sequence data. url: http://www.bioinformatics.babraham.ac.uk/projects/fastqc

Barbato, M., Orozco-terWengel, P., Tapio, M., and Bruford, M. W. (2015a). SNeP: A tool to estimate trends in recent effective population size trajectories using genome-wide SNP data. *Frontiers in Genetics*, 6, 109. doi: https://doi.org/10.3389/fgene.2015.00109

Barbato, M., Orozco-terWengel, P., Tapio, M., and M.W., B. (2015b). SNeP. url: https://sourceforge.net/projects/snepnetrends/

Berglund, P., Andonov, S., Strandberg, E., and Eriksson, S. (2024). Should performance at different race lengths be treated as genetically distinct traits in Coldblooded trotters? *J. Anim. Breed. Genet.*, 141(2), 220–234 doi: https://doi.org/10.1111/jbg.12837

Bhatia, G., Patterson, N., Sankararaman, S., and Price, A. L. (2013). Estimating and interpreting FST: The impact of rare variants. *GenomeResearch*, 23, 1514–1521. doi: https://doi.org/10.1101/gr.154831.113

Bjørnstad, G., Gunby, E., and Røed, K. H. (2000). Genetic structure of Norwegian horse breeds. *J. Anim. Breed. Genet.*, 117, 307-317. doi: https://doi.org/10.1046/j.1439-0388.2000.00264.x

Bjørnstad, G., Nilsen, N., and Røed, K. H. (2003). Genetic relationship between Mongolian and Norwegian horses? *Animal Genetics*, 34, 55-58. doi: https://doi.org/10.1046/j.1365-2052.2003.00922.x

Bjørnstad, G., and Røed, K. H. (2001). Breed demarcation and potential for breed allocation of horses assessed by microsatellite markers. *Animal Genetics*, 32, 59-65. doi: https://doi.org/10.1046/j.1365-2052.2001.00705.x

Bjørnstad, G., and Røed , K. H. (2002). Evaluation of factors affecting individual assignment precision using microsatellite data from horse breeds and simulated breed crosses. *Animal Genetics*, 33, 264–270. doi: https://doi.org/10.1046/j.1365-2052.2002.00868.x

Broad Institute. (2020). Picard Toolkit. url: https://broadinstitute.github.io/picard/

Broad Institute. (2025). Genome Analysis Toolkit (GATK). url: https://gatk.broadinstitute.org

Carlström, O., Aaby-Ericsson, A., and Wilhelmsson, F. (1946). Betänkande med förslag till åtgärder för främjande av ridhästaveln m.m. (Vol. 1946:45).(Stockholm: Kungliga boktryckeriet P.A. Nordstedt & Söner).

Chang, C. C., Chow, C. C., Tellier, L. C. A. M., Vattikuti, S., Purcell, S. M., and Lee, J. J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience*, 4(1), 7. doi: https://doi.org/10.1186/s13742-015-0047-8

Chen, C., Zhu, B., Tang, X., Chen, B., Liu, M., Gao, N., et al. (2023). Genome-Wide Assessment of Runs of Homozygosity by Whole-Genome Sequencing in Diverse Horse Breeds Worldwide. *Genes*, 14(6), 1211. doi: https://doi.org/10.3390/genes14061211

Corbin, L. J., Blott, S. C., Swinburne, J. E., Vaudin, M., Bishop, S. C., and Woolliams, J. A. (2010). Linkage disequilibrium and historical effective population size in the Thoroughbred horse. *Animal Genetics*, 41(s2), 8-15. doi: https://doi.org/10.1111/j.1365-2052.2010.02092.x

Corbin, L. J., Liu, A. Y. H., Bishop, S. C., and Woolliams, J. A. (2012). Estimation of historical effective population size using linkage disequilibria with marker data. *J. Anim. Breed. Genet.*, 129, 257-270. doi: https://doi.org/10.1111/j.1439-0388.2012.01003.x

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(2), 1-4. doi: https://doi.org/10.1093/gigascience/giab008

De Meeûs, T. (2018). Revisiting FIS, FST, Wahlund Effects, and Null Alleles. *Journal of Heredity,* 109(4), 446–456. doi: https://doi.org/10.1093/jhered/esx106

Det Norske Travselskap, and Svensk Travsport. (2019). Avelsplan för kallblods travare. url: Https://www.kallblodstravaren.se/contentassets/07906d194ac2452dad9504e90db9adfd/beslutad-svensk-norsk-avelsplan-2019-06-13.pdf

Domínguez-Viveros, J., Molina-Villalobos, J. R., Camacho-Sandoval, J., Cruz-Méndez, A., Martínez-Rocha, R., and Jahuey-Martínez, F. (2024). Structure and genetic variability of the Costa Rican Paso horse. *Journal of Equine Veterinary Science*, 132, 104985. doi: https://doi.org/10.1016/j.jevs.2023.104985

European Variation Archive (EVA). (n.d.). EMBL-EBI Catalog of equine genetic variation. url: https://www.ebi.ac.uk/eva/?eva-study=PRJEB47918

FAO. (2022). Status and Trends of Animal Genetic Resources - 2022. In 12th Session of the Intergovernmental Technical Working Group on Animal Genetic Resources for Food and Agriculture, Rome. https://www.fao.org/3/cc3705en/cc3705en.pdf

FAO. (2024). Domestic Animal Diversity Information System (DAD-IS). url: https://www.fao.org/dad-is/browse-by-country-and-species/en/

Fegraeus, K. J., Velie, B. D., Axelsson, J., Ang, R., Hamilton, N. A., Andersson, L., et al. (2018). A potential regulatory region near the EDN3 gene may control both harness racing performance and coat color variation in horses. *Physiological Reports*, 6(10), e13700. doi: https://doi.org/10.14814/phy2.13700

Föreningen Nordsvenska Hästen. (2019). Avelsprogram och reglementen för nordsvensk brukshäst. url: https://nordsvensken.org/wp-content/uploads/2019/12/AVELSPROGRAM-2019.pdf?utm_source=chatgpt.com

Frichot, E., and François, O. (2015). LEA: an R package for Landscape and Ecological Association studies. *Methods in Ecology and Evolution*, 6(8), 925-929. doi: https://doi.org/10.1111/2041-210X.12382

Gain, C., and François, O. (2021). LEA 3: Factor models in population genetics and ecological genomics with R. *Molecular Ecology Resources*, 21(8), 2738-2748. doi: https://doi.org/10.1111/1755-0998.13366

Gmel, A. I., Mikko, S., Ricard, A., Velie, B., Gerber, V., Hamilton, N. A., et al. (2024). Using high-density SNP data to unravel the origin of the Franches-Montagnes horse breed. *Genetics Selection Evolution*, 56(53). doi: https://doi.org/10.1186/s12711-024-00922-6

Gopalakrishnan, S., Samaniego Castruita, J. A., Sinding, M. H. S., Kuderna, L. F. K., Räikkönen, J., Petersen, B., et al. (2017). The wolf reference genome sequence (*Canis lupus lupus*) and its implications for *Canis* spp. population genomics. *BMC Genomics*, 18, 495. doi: https://doi.org/10.1186/s12864-017-3883-3

Greenacre, M., Groenen, P. J. F., Hastie, T., D'Enza, A. I., Markos, A., and Tuzhilina, E. (2022). Principal Component Analysis. *Nature Reviews Methods Primers*, 2, 100. doi: https://doi.org/10.1038/s43586-022-00184-w

Hästnäringens Nationella Stiftelse. (2021). Hästar och Uppfödare i Sverige. Nyckeltal för svensk hästuppfödning under åren 2016 - 2020 url: https://hastnaringen.se/app/uploads/2021/12/avelsrapport-2021-hastar-och-uppfodare-i-sverige.pdf

Hudson, R. R., Slatkint, M., and Maddison, W. P. (1992). Estimation of Levels of Gene Flow From DNA Sequence Data. *Genetics*, 132(2), 583–589. doi: https://doi.org/10.1093/genetics/132.2.583

Joensen, S. K. (2024). The Genomic Diversity and Population Structure of the Faroese Horse; The first ever whole-genome study. M.Sc. Thesis. University of Copenhagen, Copenhagen.

Kettunen, A., Joensen, S. K., and Berg, P. (2022). Optimum contribution selection (OCS) analyses prompted successful conservation actions for Faroese horse population. *Genetic Resources,* 3, 59-67. doi: https://doi.org/10.46265/genresj.KKXV5870

Kierkegaard, L. S., Groeneveld, L. F., Kettunen, A., and Berg, P. (2020). The status and need for characterization of Nordic animal genetic resources. *Acta Agric. Scand. A Anim. Sci.*, 69(1-2), 2-24. doi: https://doi.org/10.1080/09064702.2020.1722216

Kjetså, M., Gerðinum, J. I., Ólavsdóttir, J., M., J., Kallsoy Joensen, S., Honkatukia, M., et al. (2024). Action Plan for the Conservation of the Faroese Horse. url: https://www.nordgen.org/media/w4sbtasg/action-plan-for-the-conservation-of-the-faroese-horse.pdf

Landslaget for Dølahest. (2024). Avlsplan dølahest. url: https://www.dolehesten.no/avlsplan/

Leroy, G., Gicquel, E., Boettcher, P., Besbes, B., Furre, S., Fernandez, J., et al. (2020). Coancestry rate's estimate of effective population size for genetic variability monitoring. *Conserv Genet Resour*(12), 275-283. doi: https://doi.org/10.1007/s12686-019-01092-0

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics*, 25(14), 1754-1760. doi: https://doi.org/10.1093/bioinformatics/btp324

Li, Y., Liu, Y., Wang, M., Lin, X., Li, Y., Yang, T., et al. (2022). Whole-Genome Sequence Analysis Reveals the Origin of the Chakouyi Horse. *Genes*, 13(12), 2411. doi: https://doi.org/10.3390/genes13122411

Lloret-Villas, A., Bhati, M., Kadri, N. K., Fries, R., and Pausch, H. (2021). Investigating the impact of reference assembly choice on genomic analyses in a cattle breed. *BMC Genomics*, 22, 363. doi: https://doi.org/10.1186/s12864-021-07554-w

Manunza, A., Cozzi, P., Boettcher, P., Curik, I., Looft, C., Colli, L., et al. (2025). Estimating the optimal number of samples to determine the effective population size in livestock. *Frontiers in Genetics*, 16, 1588986. doi: https://doi.org/10.3389/fgene.2025.1588986

McGivney, B. A., Han, H., Corduff, L. R., Katz, L. M., Tozaki, T., MacHugh, D. E., et al. (2020). Genomic inbreeding trends, influential sire lines and selection in the global Thoroughbred horse population. *Scientific Reports*, 10, 466. doi: https://doi.org/10.1038/s41598-019-57389-5

Melheim, M. (2017). Genetisk Variasjon og Clusteranalyse på Bakgrunn av Slektskapsdata hjå Dølahest. M.Sc. thesis. Norwegian University of Life Sciences, Aas.

Meyermans, R., Gorssen, W., Buys, N., and Janssens, S. (2020). How to study runs of homozygosity using PLINK? A guide for analyzing medium density SNP data in livestock and pet species. *BMC Genomics*, 21, 94. doi: https://doi.org/10.1186/s12864-020-6463-x

Nazareno, A. G., Bemmels, J. B., Dick, C. W., and Lohmann, L. G. (2017). Minimum sample sizes for population genomics: an empirical study from an Amazonian plant species. *Molecular Ecology Resources*, 17, 1136-1147. doi: https://doi.org/10.1111/1755-0998.12654

NCBI. (2018). Genome Assembly EquCab3.0; The Equine Reference Genome. url: https://www.ncbi.nlm.nih.gov/datasets/genome/GCF_002863925.1/

Norsk Hestesenter. (2023). Nøkkeltal om dei nasjonale

hesterasane. url: https://img2.custompublish.com/getfile.php/5302809.2562.zsmbl7szjszw77/N%C3%B8kkeltallsrapport%2B2023_web.pdf?return=www.nhest.no

Olsen, H. F., Klemetsdal, G., Ruane, J., and Helfjord, T. (2010). Pedigree structure and genetic variation in the two endangered Norwegian horse breeds: Døle and Nordland/Lyngen. *Acta Agric. Scand. A Anim. Sci.*, 60(1), 13-22. doi: https://doi.org/10.1080/09064701003639884

Olsen, H. F., Tenhunen, S., Dolvik, N. I., Våge, D. I., and Klemetsdal, G. (2020). Segment-based coancestry, additive relationship and genetic variance within and between the Norwegian and the Swedish Fjord horse populations. *Acta Agric. Scand. A Anim. Sci.*, 69(1-2), 118-126. doi: https://doi.org/10.1080/09064702.2019.1711155

Petersen, J. L., Mickelson, J. R., Cleary, K. D., and McCue, M. E. (2014). The American Quarter Horse: Population Structure and Relationship to the Thoroughbred. *Journal of Heredity*, 105(2), 148–162. doi: https://doi.org/10.1093/jhered/est079

Petersen, J. L., Mickelson, J. R., Cothran, E. G., Andersson, L. S., Axelsson, J., Bailey, E., et al. (2013). Genetic Diversity in the Modern Horse Illustrated from Genome-Wide SNP Data. *PLoS One*, 8(1), e54997. doi: https://doi.org/10.1371/journal.pone.0054997

Pokharel, K., Weldenegodguad, M., Reilas, T., and Kantanen, J. (2024). EquCab_Finn: A new reference genome assembly for the domestic horse, Finnhorse. *Animal Genetics*, 33(5), 766-771. doi: https://doi.org/10.1111/age.13463

Poplin, R., Ruano-Rubio, V., DePristo, M. A., Fennell, T. J., Carneiro, M. O., Van der Auwera, G. A., et al. (2018). Scaling accurate genetic variant discovery to tens of thousands of samples. *bioRxiv*, 1-22. doi: https://doi.org/10.1101/201178

Purcell, S., and Chang, C. (2005). PLINK 1.9. url: www.cog-genomics.org/plink/1.9/

Purcell, S., and Chang, C. (2017). PLINK 2.0. url: www.cog-genomics.org/plink/2.0/

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., et al. (2007). PLINK: A tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics*, 81(3), 559-575. doi: https://doi.org/10.1086/519795

Ryman, N., Laikre, L., and Hössjer, O. (2019). Do estimates of contemporary effective population size tell us what we want to know? *Molecular Ecology*, 28(8), 1904-1918. doi: https://doi.org/10.1111/mec.15027

Schurink, A., Shrestha, M., Eriksson, S., M., B., Bovenhuis, H., Back, W., et al. (2019). The Genomic Makeup of Nine Horse Populations Sampled in the Netherlands. *Genes*, 10(6), 480. doi: https://doi.org/10.3390/genes10060480

Siekas, A.-C. (2006). Populationsstruktur och genetisk analys av exteriöra egenskaper hos svensk ardenner. M.Sc. thesis. Swedish University of Agricultural Sciences, Uppsala.

Sigurðardóttir, H., Ablondi, M., Kristjansson, T., Lindgren, G., and Eriksson, S. (2024). Genetic diversity and signatures of selection in Icelandic horses and Exmoor ponies. *BMC Genomics*, 25, 772. doi: https://doi.org/10.1186/s12864-024-10682-8

Sild, E., Rooni, K., Värv, S., Røed, K., Popov, R., Kantanen, J., et al. (2019). Genetic diversity of Estonian horse breeds and their genetic affinity to northern European and some Asian breeds. *Livestock Science*, 220, 57-66. doi: https://doi.org/10.1016/j.livsci.2018.12.006

Smogeli, N. A. (2023). Unraveling the Genetic Mysteries of the Norwegian Fjord-horse: Identifying Harmful Haplotypes for Improved Breeding Strategies. M.Sc. Thesis. Norwegian University of Life Sciences, Aas.

Stroupe, S., Millar, T., Raudsepp, T., Andersson, L., Petersen, J., Kalbfleish, T., et al. (2024). Equine pangenome graph identifies novel structural and single nucleotide variants. In 14th International Havemeyer Foundation Horse Genome Workshop, Caen. https://www.pure.ed.ac.uk/ws/portalfiles/portal/443373075/Abstracts-book-040524.pdf

Svensk Travsport. (2024). Kallblodstravarnas betäckningssiffror för 2024. url: https://www.travsport.se/arkiv/nyheter/2024/oktober/kallblodstravarnas-betackningssiffror-for-2024/

Svenska Hästavelsförbundet. (2024). Betäckningssiffor 2024. url: https://svehast.se/wp-content/uploads/2024/11/Betackningssiffror_2024.pdf.

Svenska Russavelsföreningen. (2019). Svenska Russavelsföreningens avelsprogram samt avelsplan för avel med gotlandsruss 2020. url: https://usercontent.one/wp/www.gotlandsruss.se/wp-content/uploads/2020/02/2020-Svenska-Russavelsf%C3%B6reningens-avelsprogram-samt-avelsplan-f%C3%B6r-avel-med-gotlandsruss.pdf

Thorburn, D.-M., Sagonas, K., Binzer-Panchal, M., Chain, F. J. J., Feulner, P. G. D., Bornberg-Bauer, E., et al. (2023). Origin matters: Using a local reference genome improves measures in population genomics. *Molecular Ecology Resources,* 23(7), 1706–1723. doi: https://doi.org/10.1111/1755-0998.13838

Van der Auwera, G. A., and O'Connor, B. D. (2020). Genomics in the Cloud: Using Docker, GATK, and WDL in Terra (1st Edition).(Sebastopol, California: O'Reilly Media).

Velie, B. D., Lillie, M., Fegraeus, K. J., Rosengren, M. K., Solé, M., Wiklund, M., et al. (2019a). Exploring the genetics of trotting racing ability in horses using a unique Nordic horse model. *BMC Genomics,* 20, 104 (2019). doi: https://doi.org/10.1186/s12864-019-5484-9

Velie, B. D., Solé, M., Fegraeus, K. J., Rosengren, M. K., Røed, K. H., Ihler, C.-F., et al. (2019b). Genomic measures of inbreeding in the Norwegian-Swedish Coldblooded Trotter and their associations with known QTL for reproduction and health traits. *Genetics Selection Evolution*, 51, 22. doi: https://doi.org/10.1186/s12711-019-0465-7

Viklund, Å., Näsholm, A., Strandberg, E., and Philipsson, J. (2011). Genetic trends for performance of Swedish Warmblood horses. Livestock Science*, 141(2-3), 113-122.* doi: https://doi.org/10.1016/j.livsci.2011.05.006

Weldenegodguad, M., Popov, R., Pokharel, K., Ammosov, I., Ming, Y., Ivanova, Z., et al. (2019). Whole-Genome Sequencing of Three Native Cattle Breeds Originating From the Northernmost Cattle Farming Regions. *Frontiers in Genetics*, 9, 728. doi: https://doi.org/10.3389/fgene.2018.00728

White, E.-L. F., Honkatukia, M., Peippo, J., and Kjetså, M. (2024). Equines in the Nordics: History, Status and Genetics. url: https://www.norden.org/en/publication/equines-nordics-history-status-and-genetics

Wright, S. (1922). Coefficients of inbreeding and relationship. *American Naturalist*, 56(645), 330–338. url: http://www.jstor.org/stable/2456273

Zhdanova, O. L., and Pudovkin, A. I. (2008). Nb_HetEx: A Program to Estimate the Effective Number of Breeders. *Journal of Heredity*, 99(6), 694–695. doi: https://doi.org/10.1093/jhered/esn061

Genetic Resources

ORIGINAL ARTICLE

Genetic Resources (2026), 7 (13), 118–128
DOI: 10.46265/genresj.GPKT7328
https://www.genresj.org
ISSN: 2708-3764

A multidisciplinary framework for adding value to the indigenous cattle breed of Cyprus

Anna Spyrou[a,*], Andreas C. Dimitriou[b,d], Valeria Mattiangeli[c], Deborah Diquelou[c], Daniel G. Bradley[c], Victoria E. Mullin[c], Georgia Hadjipavlou[d,*]

[a] *Science and Technology in Archaeology and Culture Research Center (STARC), The Cyprus Institute, 20 Konstantinou Kavafi Street 2121, Aglantzia, Nicosia*
[b] *Department of Veterinary Medicine, School of Veterinary Medicine, European University Cyprus, 2404, Engomi, Nicosia, Cyprus*
[c] *The Smurfit Institute of Genetics, School of Genetics and Microbiology, Trinity College Dublin, Dublin 2, D02 VF25, Ireland*
[d] *Department of Animal Production, Agricultural Research Institute, 1516 Nicosia, Cyprus*

Abstract: The officially recognized Cyprus cattle breed is the island's only native cattle breed, having evolved and adapted over millennia to local environmental conditions and rural traditions. During pre-industrial times, the island's indigenous cattle were exclusively used in agricultural production and for the transportation of people and goods. Beyond their utilitarian roles, cattle held a special position in the island's cultural sphere, as reflected in iconography, food taboos, human-cattle cohabitation, and the integration of cattle into major religious celebrations. The mechanization of agriculture along with the increasing demand for high-yielding breeds, followed by large-scale urban development, resulted in the sharp decline of the breed's population, leading to its current endangered status. 'Animals REsilient in TIme' (ARETI) is a research project that aimed to decipher the economic, cultural and genetic history of the island's indigenous cattle breed from antiquity to the present. Through a multidisciplinary research framework that integrates evidence from zooarchaeology, history, genomics, ethnography and folklore studies, the project deepened our understanding of the breed's unique genetic traits and long-standing connection to people and local environments, strengthening the prospects for its valorization, conservation and sustainable use. In this article, we outline the project's multidisciplinary framework and propose that similar approaches could be extended to support the conservation of other endangered animal breeds in Cyprus, the Eastern Mediterranean and beyond.

Keywords: Animal Genetic Resources (AnGR), Cyprus indigenous cattle breed, zooarchaeology, genetic characterization, cattle palaeogenomics, *Bos taurus-Bos indicus* hybrid

Citation: Spyrou, A., Dimitriou, A. C., Mattiangeli, V., Diquelou, D., Bradley, D. G., Mullin, V. E. and Hadjipavlou, G. (2026) "A multidisciplinary framework for adding value to the indigenous cattle breed of Cyprus", *Genetic Resources*, 7(13), pp. 118–128 doi: 10.46265/genresj.GPKT7328

Introduction

Due to its proximity to the Middle East, the cradle of plant and animal domestication, the Mediterranean region is among the world's richest in livestock breed diversity, harbouring animal genetic resources (AnGR) that have evolved during the last 10,000 years through a combination of natural selection, traditional farming practices and complex social and cultural processes (Boyazoglou and Hatziminaoglou, 2002; Ligda *et al*, 2022). As in many other local breeds, this complex evolutionary history, along with periods of genetic isolation, has led to the development of numerous distinct animal breeds with unique phenotypic traits and genetic structures (Hall, 2019). Local or indigenous livestock breeds play a crucial role in sustainable rural livelihoods and the utilization of marginal ecological areas (Köhler-Rollefson, 2003). Besides providing a wide variety of products, these breeds are well adapted to extreme environmental conditions such as high temperatures, limited water availability and low-quality forage. They also exhibit notable resistance to diseases, making them a vital genetic reservoir for addressing

* Corresponding authors: Anna Spyrou (a.spyrou@cyi.ac.cy), Georgia Hadjipavlou (ghadjipavlou@ari.moa.gov.cy)

Received: 14.08.2025 | Accepted: 10.02.2026 | Published online: 23.03.2026

current and future challenges related to environmental stress and climate change (Mendelsohn, 2003; da Gama, 2006; Hadjipavlou and Ligda, 2022). The increasing frequency of wildfires driven by climate change, particularly in the region, further underscores the valuable potential of these livestock breeds. Through their natural grazing behaviour, they can help reduce wildfire risk and/or mitigate the extent of damage caused by such events (Pillar and Overbeck, 2025).

In recent decades, AnGR in the Mediterranean have been experiencing an alarming rate of erosion due to a range of socioeconomic factors, including the intensification of agricultural practices, after the Second World War, economic globalization, the widespread use of a few highly productive breeds, large-scale urbanization and the abandonment of rural areas (Rischkowsky and Pilling, 2007; Gandini *et al*, 2010). The loss of native animal breeds not only reduces biodiversity per se but also leads to the reduction of valuable genetic variability essential for future breeding programmes (Hoffman, 2010). However, the importance of AnGR extends beyond biology and genetics to encompass the symbolic and material ways in which they are connected to humans and regional environments (Ovaska *et al*, 2021). Pictorial records from dynastic Egypt suggest the existence of different dog (Thiringer, 2020) and cattle breeds (Redding, 2024) as early as the 4th millennium BCE, demonstrating that the development of domestic animal breed diversity is a long-standing process with roots in early prehistory (Köhler-Rollefson, 1997). Thus, domestic animal breeds are also repositories of cultural heritage, embodying centuries of human–animal coevolution and interaction and serving as living 'monuments' of local traditions and identities (Hall, 2019). As animals are deeply embedded in agricultural traditions and cultural practices (Gandini and Villa, 2003) and often linked to local myths and narratives (Haudricourt, 1962), their population decline threatens the continuity of long-standing husbandry systems, oral traditions, ecological knowledge and the collective cultural memory of rural communities.

Like many other indigenous cattle breeds in the Mediterranean region, the indigenous cattle breed of Cyprus has experienced a significant population size decline over the past century, leading to its current endangered status. According to the most recent census conducted by the Cyprus Ministry of Agriculture, Rural Development and the Environment, there are approximately 1,304 native cattle remaining on the island, of which only 709 are breeding females (DAG Report, 2024). These population figures place the indigenous cattle breed of Cyprus on the FAO's list of 'endangered farm animal breeds' (Scherf, 2000). Cattle were introduced to Cyprus during the 9th millennium BCE, shortly after their domestication in the Fertile Crescent, more specifically, in northern Syria (Vigne *et al*, 2014). Since Cyprus is a truly oceanic island and has not been connected to the nearby mainland since at least the end of the Miocene (Dimitriou *et al* (2022), and references therein), the transportation of cattle to Cyprus must have occurred via sea, likely on dugout canoes (Vigne *et al*, 2014). Cyprus is the only island in the Eastern Mediterranean known to have experienced the disappearance of cattle during the Neolithic for reasons that remain unclear (Davis, 2003; Horwitz *et al*, 2004). After an absence of nearly three millennia, cattle were reintroduced to Cyprus at the dawn of the Early Bronze Age (mid-third millennium BCE) by migrant populations from Anatolia and Cilicia. From the third millennium BCE to the present day, cattle have maintained a continuous presence on the island. The fate of the contemporary indigenous cattle breed, particularly during the 20th century, has been interwoven with economic, social, political, environmental and other issues that transformed the relationship between people and production systems, as well as between people and the environment. Significant declines in the population of the breed have been recorded between the years 1960s and 2008, with numbers dropping by 97% during this period (Dimitriou et al, 2024a). The most significant shifts, however, occurred between 1965–1970 and 1973–1974, linked respectively to the introduction of high-yielding breeds (e.g. Holstein Friesian) and the abandonment of rural areas following the 1974 invasion. The breed's population reached its lowest point in 2008, with only 746 individuals remaining (DAG Report, 2024).

Over the past decade, the population size of the Cyprus local cattle breed has increased and remained stable, thanks to the combined efforts of government subsidy programmes further supported by the European Union since 2004, and the Breeders' Association, established in 2010 and recognized as a Breeder Society in 2020. The breed is sporadically utilized for its ecosystem services, particularly in controlling reed expansion through traditional grazing practices. Its ecological value was recently highlighted by a Darwin Plus project, which demonstrated the breed's effectiveness in managing the biodiversity-rich wetlands of the Akrotiri Peninsula and its contribution to maintaining a diverse mosaic of habitats (Vayanou *et al*, 2024). Additionally, small-scale meat production involving the breed is carried out by individual breeders, though it lacks branding and marketing efforts. Meanwhile, plans are underway to position meat from the indigenous cattle as a delicacy product targeted at niche markets, which may create new incentives for conservation through sustainable use (Ligda and Casabianca, 2013).

Despite the island's narrow geographic limits and the relatively small population size of its livestock, the genetic resources of local breeds remain largely unexplored. Exceptions include the recently published comprehensive study on the Cyprus Chios sheep (Dimitriou et al, 2024b) and data representing only a few individuals regarding the local cattle breed. More specifically, genetic data from animals representing the Cyprus local cattle breed were generated within the framework of studies aiming at exploring the genetic diversity of locally adapted breeds across broader geographic and environmental scales (Flori *et al*, 2019; Papachristou *et al*, 2020; Papachristou, 2023). Yet, their scope did not allow for a comprehensive investigation of the focal breed, as they did not include an adequate number of individuals representing the full extent of the taxon's distribution.

In 2022, a team composed by a zooarchaeologist, two animal biologists as well as geneticists and palaeogeneticists embarked on a multidisciplinary research project to decipher the genetic, economic and cultural history of the indigenous cattle breed of Cyprus through an innovative research methodology that blended cutting-edge scientific techniques with anthropologically oriented approaches. By weaving together evidence from zooarchaeology, genomics, ethnography and folklore studies, the project, known by the acronym ARETI (Animals REsilient in TIme) provided an opportunity to unearth vital information about the past, present and future of cattle in Cyprus and their interaction with human societies from antiquity to the present.

Materials and methods

The project has been developed around six key pillars, which also provided its methodological framework. These are briefly discussed below:

1. **Zooarchaeological research:** Zooarchaeological data were extracted and compiled from existing literature, spanning from the Pre-Pottery Neolithic B (PPNB) to the Byzantine period, including faunal material from four key archaeological sites on the island. In addition to literature review, the same cattle bone assemblages were re-examined for taphonomic modifications to unravel patterns of human–cattle interaction. Key variables on the bones of cattle were examined, including age-at-death profiles, cut marks, pathological indicators, and other skeletal modifications, all of which offer valuable insights into the roles cattle played in subsistence economy, labour and ritual practices over time and across space.
2. **Ancient DNA analysis:** To investigate the extent of *Bos indicus* introgression in ancient Cypriot cattle, we conducted ancient DNA (aDNA) analyses on a small sample of cattle bones (n = 16) recovered from the same assemblages examined under Pillar 1. DNA was extracted in clean-room facilities dedicated to ancient DNA research at Trinity College Dublin and was carried out following a combined protocol (Gamba *et al*, 2014; Boessenkool *et al*, 2017; Dabney *et al*, 2019). DNA extracts were treated with USER enzyme to reduce post-mortem deamination lesions, and double-stranded libraries were created for Illumina sequencing (Meyer and Kircher, 2010). Pair-end sequencing was carried out on a NovaSeq 6000 platform (S1, 100bp) in TrinSeq, Dublin.
3. **Modern DNA analysis**: More than 120 animals from the extant local Cyprus cattle breed were sampled for DNA extraction and genotyping. All animals originated from 15 farms belonging to members of the local Cyprus Cattle Breeder Society and were part of the Breeder Society herdbooks (i.e. retained the basic phenotypic characteristics of animals belonging to the breed, even though no actual phenotyping recording took place within the current project). Total genomic DNA was isolated from all samples, while DNA extractions from Greece and Turkey were requested and kindly provided to us by colleagues. The inclusion of individuals representing neighbouring Greek island and Turkish cattle populations aimed to further investigate the postulated genetic connection of these populations with the Cyprus cattle breed (Flori *et al*, 2019; Papachristou *et al*, 2020). Genome-wide single nucleotide polymorphism (SNP) data for all available specimens were generated using the Illumina 777K BovineHD BeadChip. The final dataset was further enriched with genotypes for individuals from commercial breeds that are currently, or were historically, present on the island (i.e. Limousin, Charolais, Jersey, Holstein and Angus), with the aim of exploring possible gene flow between the focal breed and these latter populations. Non-model-based analyses, such as principal component analysis (PCA), as well as model-based methods including population structure and phylogenetic analyses, were employed to detect possible divergence within the Cypriot population and to assess the genetic influence of the local neighbouring and commercial breeds on the genetic profile of the Cyprus cattle. Finally, inbreeding indices and effective population size were calculated as indicators of population health and as tools for management planning.
4. **Archival research:** Archival research was central to the project, enabling us to reconstruct the breed's recent history and to contextualize how Cyprus' indigenous cattle were perceived, bred and managed within broader historical, economic, sociocultural and political frameworks. We consulted a wide range of materials held at the Cyprus State Archives, including annual reports by Directors and other officials of the Department of Agriculture, agricultural censuses, property and land surveys, travellers' accounts and memoirs, British colonial reports, historical magazines and agricultural journals. Our review focused primarily on the period from the beginning of British colonial rule (1878) to the early 1980s, allowing us to trace the breed's trajectory across a pivotal era in Cyprus' modern history. We also consulted selected historical sources relating to the island's Lusignan era, including de Mas Latrie's Histoire de l'île de Chypre (de Mas Latrie,1855).
5. **Ethnographic research:** Ethnographic research was essential to the project for understanding the perceptions, practices and broader cultural frameworks through which traditional Cypriot society integrated animals into everyday life. The ethnographic component combined (1) a review of relevant ethnographic and historical literature with (2) primary qualitative data collection among cattle breeders and older farmers. Primary data were gathered through unstructured interviews, informal conversations (including discussions with cattle breeders in village coffee shops) and focus groups, alongside the collection of oral histories, including collective/group oral history sessions. These complementary methods enabled the project to explore the diverse ways in which cattle were embedded in the social life of the island's inhabitants, including food taboos related to the consumption of cattle meat and the avoidance of milk – particularly among pregnant women (Ohnefalsch-Richter, 1913) – naming practices, human–cattle cohabitation (Panaretos 1967; Xioutas 1978), and the involvement of animals in the veneration of saints and major celebrations of the Christian Orthodox faith (Neophytou, 2014). In addition, the project examined the symbolic and imaginative roles of animals as reflected in oral traditions, folktales, songs and proverbs.
6. **Public engagement:** An essential requirement for the conservation of native animal breeds is broad recognition and understanding of their importance within society (Mendelsohn, 2003). As has been recently emphasized, conservation strategies are more effective when initiated at the community level and employ participatory methods, rather than relying solely on top-down approaches (Soini *et al*, 2012). For the ARETI project, this has been implemented through various initiatives, including informative seminars for the breeders, interactive activities for schoolchildren, collaboration with local artists, and the production of a documentary film. The involvement of the Breeder Society at various stages of the project proved invaluable, ensuring their voices were heard and helping them recognize the importance of their role as custodians of genetic diversity.

In the sections that follow, we outline the main results of each pillar, highlighting how they shed light on different aspects of the breed's ancient and historical background, while also contributing to its current valorization in both academic and public spheres.

Results

Zooarchaeological research

The Zooarchaeological work conducted during this project shed light on several interesting aspects of human-cattle interaction. A particularly notable practice in Late Bronze–Early Iron Age Cyprus, which is also attested in nearby regions (van Dijk, 2013) is linked to the modification of cattle skulls for the making of bucrania, reflecting the symbolic and potentially ritualistic significance of cattle in ancient Cyprus and the Eastern Mediterranean World (Averett, 2020). An interesting observation in the zooarchaeological assemblages dating to the Early-Middle Bronze Ages is the general absence of cut marks, in contrast to the frequent cut marks observed on the remains of smaller livestock species, particularly domestic caprine and pig. The absence of butchery marks from the bones of cattle may reflect a cultural choice, potentially linked to the animals' primary role in agriculture, to religious or symbolic significance, or to a combination of both factors. In addition, age-at-death data suggests a management strategy in which cattle were kept until old age, likely due to their prolonged use in agricultural labour (Croft, 2006). This is further supported by the identification of pathologies (e.g. osteoarthritis) on cattle phalanges, suggesting that cattle were subjected to mechanical pressure during their use in agricultural activities (Thomas and Johanssen, 2011). However, a different pattern appears to emerge toward the end of the Late Bronze Age, with cut marks becoming more frequent and visible on cattle bones, potentially linked to urbanization and an increased demand for meat. Beyond their economic role, cattle may also have functioned as social capital, featuring in rites of passage such as funerary practices and bridewealth transactions. Additional insights into the role of cattle in prehistoric Cyprus have been gained through stable isotope analysis (Spyrou *et al*, 2024), which indicates that cattle were subject to more intensive human management than smaller livestock. This included practices such as restricted grazing and possible provisioning with fodder, self-fertilized hay, or silage. Combined with iconographic evidence, these findings suggest that cattle and oxen were valued more highly than other animals. Overall, the results from the zooarchaeological study help trace the deep historical roots of human–cattle relationships in Cyprus, highlighting the significant role of cattle and oxen in the island's ancient economy, society and culture. These results will be published soon in a separate publication.

Cattle palaeogenomics

Due to the island's environmental conditions, only one sample preserved sufficient genetic material to allow further investigation into the animal's genetic identity (~6% endogenous content, coverage 0.02X). This sample, a petrous portion of the temporal bone from a male cow, was recovered from the monumental sanctuary at Kition Kathari (Karageorghis, 2005). Its earliest construction phase dates to the end of the second millennium BCE. Based on its archaeological context, the sample is dated to the 8th–7th century BCE, corresponding to the Early Cypro-Archaic period of the Iron Age. Mitochondrial DNA analysis (5.6X coverage, 20% missingness) indicates that the maternal line of the sample from ancient Kition belongs to haplogroup T3, one of the five major mitochondrial haplotype clusters in *Bos taurus* (T*, T1, T2, T3, T4). Haplogroup T3 is commonly found in taurine cattle populations across Europe and Southwest Asia today (Achilli *et al*, 2009). On the other hand, whole genomic analysis (admixture and D-statistics (Pritchard *et al*, 2000; Green *et al*, 2010)), where sample genetic variability was compared to both ancient and modern samples belonging to both *Bos taurus* and *Bos indicus*, revealed that the animal was a *Bos taurus–Bos indicus* hybrid (Figure 1). Admixed animals became common in Southwest Asia from the third millennium BCE onwards (Verdugo *et al*, 2019). The significance of this finding is discussed further below (Discussion).

Unfortunately, no genetic samples dating to the Early Neolithic and Early/Middle Bronze Age, when the new population of cattle was introduced to the island after three millennia of absence, have been recovered to date. It is possible that the hybrid animal identified at ancient Kition was introduced to Cyprus long before the first millennium BCE, but in the absence of genetic evidence, this remains a hypothesis.

Genetic characterization of the contemporary indigenous cattle population

Our preliminary findings showed that the Cyprus local cattle breed is genetically distinct from neighbouring indigenous breeds. Also, no significant gene flow was detected between commercial cow breeds and the Cyprus local cattle breed. Initial findings indicate the presence of two genetic subpopulations within Cyprus. However, further research is required to verify this finding and its potential connection with the previously reported geographic pattern distribution of the two morphotypes. With respect to the genetic findings relevant to conservation efforts, relatively high inbreeding was detected (mean FROH > 0.1), as expected given the major genetic bottleneck the breed had suffered in 2008, and indicating the need for better reproductive management to avoid mating highly related individuals. In addition, the effective population size appears to follow a declining trajectory, estimated at 112 individuals about 13 generations ago, indicating a constant loss of genetic variation associated with population size reduction. A publication compiling all modern DNA findings is currently in preparation.

Archival research

The wealth of archival data collected has allowed the project to establish the breed's significance over time, understand breeding and management schemes, track past population sizes, and identify distribution trends and changes. According to historical sources, Cyprus had approximately 55,000 cattle during the Lusignan period (12th–15th centuries CE) (de Mas Latrie, 1852). This figure highlights the vital role of bovine labour in threshing, ploughing, and the transport of people and goods in medieval Cyprus. During the British period, the island's cattle population was recorded at 28,919 in 1881 (The Cyprus Blue Book 1881), equivalent to approximately one animal for every five people living in rural areas

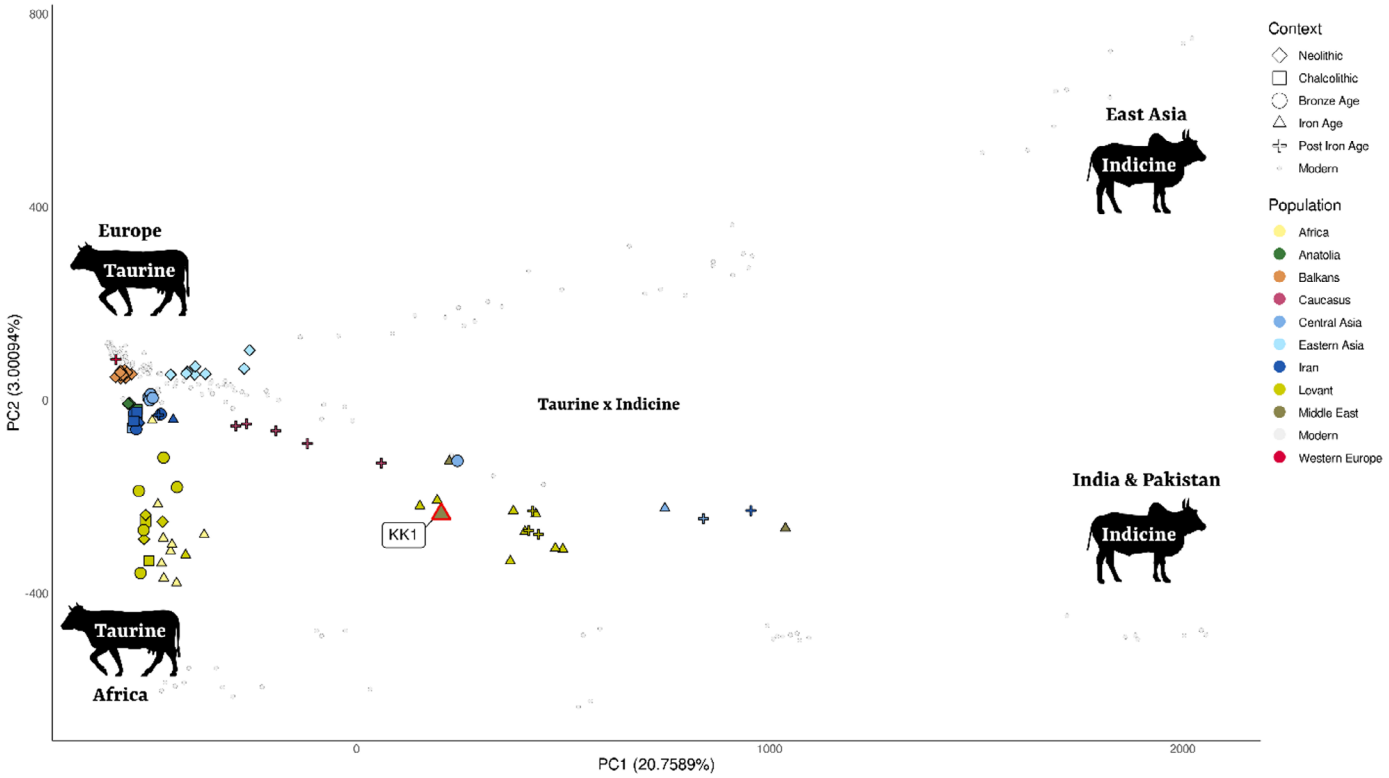


Figure 1. Projection principal component analysis (PCA) analyses of cattle; ancient cattle are projected onto modern cattle diversity. The KK1 sample falls in the hybrid *Bos taurus–Bos indicus* zone. (Adapted from Verdugo *et al*, 2019)

Figure 2. Using the traditional plough in mountainous Cyprus (1958). With the kind permission of the Cyprus Press and Information Office

(Rizopoulou-Egoumenidou 2009). These figures refer to the island's native cattle, as the first dairy cattle were introduced only in 1912 (Constantinides, 1955). Furthermore, archival research has helped pinpoint key moments of decline, such as the mass slaughter and export of oxen in Egypt in 1870 due to a drought that severely affected agriculture (Savile, 1878). Additionally, historical photographs sourced from the Photographic Archive of the Cyprus Press and Information Office have played a key role in unravelling the recent history of the indigenous cattle breed of Cyprus. Beyond their aesthetic value, these images provided crucial historical context, as many depict the animals in association with official, political figures or significant historical events. Most importantly, they depict the animals in association with the Cyprus rural environment, providing a rather nostalgic image of a 'lost' past (Figure 2).

Archives played a dynamic role in the project that went beyond documentation, enrichment and research. Some of the older photographs were used as 'triggers' in discussions with local breeders, facilitating conversations about ancient and notably 'primitive' phenotypic traits of the breed, such as the distinctive white ring around some of the animals' noses. Historical images depicting the breed engaged in agricultural tasks or interacting with humans in Cyprus's rural landscape also proved to be powerful memory triggers, particularly for older breeders, who recalled key agricultural practices, customs, festivals and religious events in which the indigenous cattle breed had a strong participation. Furthermore, to make this material accessible to a wider audience, and to use it as a 'living archive' (Sabiescu, 2020) through which people can remember, reflect and (re) create new spaces for collective memory, the project is developing a documentary that will present the archival content to the broader public, encouraging a dialogue and allowing multiple layers connected to the breed to be visible.

Ethnographies of human-cattle interaction and collection of oral history

In pre-industrial Cyprus, cattle and oxen were integral to bride wealth transactions (personal communication with elderly breeders) and played a prominent role in major events of the Christian Orthodox religious calendar (Neophytou, 2014). On New Year's Day, the villagers shared the joy and blessings of their faith with their most loyal companions in life – their oxen. They would light up the stable, rejoice with great delight, and consider it a good omen for the first being to enter their home that day to be an ox. They chose their finest ox and brought it into the house (Panaretos 1967; Xioutas 1978; Neophytou, 2014). They offered the animals koliva, a liturgical Eastern Orthodox dish made of barley, pomegranate seeds and almonds, and recited the following rhyme: "Eat, so we may eat from our common labour." Another custom associated with New Year's Eve or Epiphany in some villages involves placing lit candles on the horns of oxen. This ritual was believed to bring the animals a holy blessing, ensuring their health and prosperity during the agricultural season (Panaretos, 1967; Xioutas, 1978).

From the very beginning of the project, interviews have been conducted with the cattle breeders, aiming to understand their perspective relevant to the conservation of the breed. Several questions were raised, including the main motivations for keeping local cattle, their interaction with these animals, their main concerns as well as how they envision the future of the breed. Beyond financial support through subsidies, which most cattle breeders consider insufficient, the primary motivation for maintaining native cattle is deeply rooted in tradition, culture and a strong emotional connection between breeders and the animals. When asked how they would feel about the potential loss of the breed, most breeders referred to broader cultural losses, including the disappearance of vernacular dialect relating to traditional agriculture, as well as the erosion of our collective past that links us to the land and the animals. Many continue to raise native cattle because their fathers and grandfathers did so, expressing that it would be a profound loss for society to witness the extinction of an animal breed that supported so many generations of people on the island. The special bond between humans and their oxen is also reflected in the traditional avoidance of beef – a dietary taboo that persisted until the second half of the 20th century, when the importation of exotic beef breeds became more widespread on the island (Scott-Stevenson, 1880; Ohnefalsch-Richter, 1913). In some Cypriot villages, this avoidance is still observed. According to Panagiotis Gennadius, the first Director of the Department of Agriculture in Cyprus, "Almost all peasants considered the killing of a cow or an ox an act of sacrilege (Apostolou, 2018)." Similarly, it was considered a sin to burn old ploughs.

Another interesting aspect of this profound connection is the close cohabitation between humans and their oxen. In the 1930s, Brewster J. Surridge highlighted that during working times oxen were fed through the night and "[…] are kept in the room where their owner sleeps. They give warmth during the cold weather, and they are generally considered to be a source of health" (Surridge, 1930). The co-habitation aspect has also been raised by elderly breeders. In addition, the practice of naming animals based on their external characteristics demonstrates the emotional bond between humans and animals in small-scale farming societies. Today, cattle breeders often name their animals after the breeder from whom they received them, highlighting the social networks involved in the exchange of animals.

The role of cattle and oxen in the island's oral traditions, including songs, myths, folktales, proverbs and folk-art paintings, constitutes another important archive of collective memory, which needs to be safeguarded and transmitted from generation to generation. Most proverbs about oxen are closely tied to agriculture and the figure of the farmer, consistently emphasizing both the immense value of oxen as working animals and the hardship of their labour, especially threshing, which is regarded as one of the most strenuous tasks for these animals. A particularly interesting proverb that reflects this reality is: "Would the young man not harvest, would the young woman not give birth, and would the ox not thresh, none of them would grow old" (Panaretos, 1967; Xioutas, 1978). Notably, a folktale about the ox and the ant (Kleridis, 2017) makes special reference to the hump of Cyprus's indigenous cattle, an observation that aligns with previous genetic studies indicating a high level of indicine ancestry (Papachristou, 2023). Ethnographic accounts and oral traditions assign deep cultural meaning to the island's indigenous cattle, which were so deeply woven into everyday life that their presence in songs, myths and proverbs is both frequent and meaningful. This reinforces the notion that conserving an indigenous breed involves more than preserving its genetic material; it also requires safeguarding the traditional livestock systems, oral histories and cultural practices associated with it. Breeders may continue to keep native animals, but without these elements, the genetic value will be reduced (Hall, 2019).

Public engagement and citizen science initiatives

Aiming to reach out to the island's wider community, the team prepared a documentary film that raises awareness of the long-standing history, cultural and ecological values of the island's native cattle and brings to the forefront the main threats that the animals are currently facing. The involvement of different stakeholders – including breeders, researchers, public authorities, older generation farmers and NGOs – allowed multiple voices to be heard, providing a fruitful

discourse concerning the urge for conservation. Beyond informing and raising awareness, the documentary can be understood as another, non-physical 'space' for preserving the breed and the collective memory surrounding it, helping to ensure that the many stories, oral traditions, and customs associated with it are safeguarded and passed down through generations.

As part of efforts to promote citizen science (e.g. Fraisl *et al*, 2022), a collaboration has been established with local artists to produce handcrafted candle figurines inspired by the history of the breed. Each figurine has been packaged in a box and accompanied by a small leaflet, prepared in collaboration with the breeders, featuring the cow's name, a brief profile of the specific animal, and a miniature ear tag modelled after those used on living cows. The figurines were available during the screenings and fairs, with proceeds going to the Breeder Society to support ongoing conservation efforts. With each purchase, citizens were symbolically 'adopting a local cow,' actively contributing to the preservation of the indigenous cattle breed of Cyprus. Finally, the project's long-term vision is to ensure that research, activism and education concerning the indigenous cattle breed of Cyprus will have a positive impact on the animals and the environment.

Discussion

Cattle were among the first wild terrestrial species meeting the conditions for successful domestication by humans, including an herbivorous diet, rapid growth, the ability to breed in captivity, a genetic predisposition toward reduced aggression in enclosed settings, and social behaviours that facilitated handling (Felius *et al*, 2014). The wide diversity of different cattle breeds, existing today in the world, is the product of biological processes and human interventions – including sociocultural breeding regimes and economic utilization patterns – that were initiated about 10,000 years ago in the Fertile Crescent (Köhler-Rollefson, 2001). As such, the huge diversity of cattle breeds reflects not only ecological but also cultural diversity or, as Köhler-Rollefson and Meyer (2014) have argued, these breeds reflect a form of biocultural heritage that needs to be safeguarded.

This article presents an inclusive framework for studying the indigenous cattle breed of Cyprus, stemming from a multidisciplinary approach that combines the humanities and life sciences. It also extends to other endangered animal breeds through the integration of diverse concepts and methodologies developed during the ARETI project. Like many other local breeds in the Mediterranean region, Cyprus indigenous cattle remain insufficiently acknowledged and valued within the island's environmental, cultural and commercial contexts. To support the breed's valorization in both academic and public domains and to further promote its conservation, the ARETI project employed a multidisciplinary research methodology, structured around six key pillars. These pillars revealed various dimensions of the breed's ancient and historical presence, genetic value and legacy, as well as its socio-cultural significance, reinforcing its importance within Cyprus ecological and cultural landscape and further supporting its sustainable use. Zooarchaeology and palaeogenomics provided insights into past human–cattle interactions, including patterns of animal translocation in antiquity, and highlighted the deeply rooted tradition of cattle-keeping on the island. The identification of a *Bos taurus–Bos indicus* hybrid, dating to the first millennium BCE, not only supports earlier hypotheses been solely based on iconography (Spyrou, 2021; Figure 3) but also places Cyprus in the wider social and cultural networks during the Late Bronze-Early Iron Age. This genomic finding further highlights the zooarchaeological findings on the sociocultural value of cattle in Cyprus by linking it to historical processes of human mobility, animal exchange, adaptation, and human–animal interactions on the island. Indicine introgression into the Near Eastern cattle populations is estimated to have begun approximately 4,000 years ago (Verdugo *et al*, 2019). The fact that this gene flow was male driven suggests deliberate human selection (Verdugo *et al*, 2019), reflecting patterns of biological translocation across the region during the Bronze Age, alongside the exchange and trade of other key elements of material culture, such as copper, a main historical product of Cyprus. Human selection of *Bos indicus* males for breeding is thought to have been a strategic response by ancient farmers to the high temperatures affecting the region, including the island of Cyprus, at the end of the third millennium BCE (Kaniewski *et al*, 2019).

The genetic characterization of many individuals of the remaining contemporary population in this study demonstrated the breed's genetic differentiation from exotic breeds on the island and from local cattle breeds in the eastern Mediterranean area. Most importantly, the highlighted genetic uniqueness of the local breed could serve as a strong incentive for the Breeder Society and local authorities to actively support efforts aimed at preserving the local population. Until recently, limited genetic data were available for Cyprus local cattle individuals, generated within the framework of comparative studies with other breeds (Flori *et al*, 2019; Papachristou *et al*, 2020; Papachristou, 2023). These studies reached different conclusions regarding the genetic affinity of Cyprus cattle with neighbouring breeds. More specifically, the results of Papachristou *et al* (2020) supported the grouping of the Cyprus population with individuals from other breeds in Greece. Morphologically, this appears to be a reasonable result, considering the similarities between local breeds in

Figure 3. Terracotta figurine of a zebu (*Bos indicus*) from Ayia Irini, LCII-III, Cyprus Museum, Nicosia (Inv. No. 1984/1-21/1).

the southern and eastern Greek islands. However, this is not the case when comparing animals from Cyprus and northern Greece. On the other hand, the findings of Flori *et al* (2019) indicated a closer relationship between Cyprus individuals and individuals from breeds in Turkey. This outcome was also not surprising, given the subtle morphological differences between Cyprus local cattle and Turkey's local breeds, such as the Anatolian Southern Yellow. We assume that these contrasting results could be attributed to the composition of the analyzed datasets, including, on one hand, populations from various/different neighbouring breeds and on the other hand, a limited number of individuals from Cyprus, probably not adequately representing the local genetic diversity. These results need to be further confirmed using a larger sample size.

The collection of archival material provided the socio-historical context for the research, while the ethnographic data offered insights into how the traditional Cypriot society integrated cattle and oxen into both everyday life and the island's broader cultural landscape. The establishment of a dialogue with the Breeder Society proved especially valuable to the project not only because it facilitated access to samples for the breed's genetic characterization, but also because breeders have shared many stories that helped to better understand livestock transaction systems, breeding techniques and local concepts of animal care.

Moreover, public engagement played a vital role in disseminating the project's findings to the wider community. The production of a documentary has further supported this goal, serving as both a repository and a non-physical space for preserving the breed and its unique biocultural heritage. Finally, the project's multidisciplinary and collaborative approach enabled team members to contribute their specialized knowledge, share insights, and collectively interpret their findings within a broader context, demonstrating the strength of integrated interdisciplinary research in addressing complex conservation challenges.

Given the breed's significant genetic, historical and sociocultural value, as demonstrated through this project, an in situ conservation strategy should be prioritized, with particular emphasis on preserving the intangible assets of the breed along with the conservation of its genetic diversity. The need for a better and more sustainable utilization of the breed within the island's ecological and cultural landscape is essential. This includes greater utilization of the breed's ecosystem services, as already demonstrated through its role in biodiversity maintenance at the Akrotiri Marshes. Moreover, targeted grazing by indigenous cattle could help reduce flammable biomass in forest-adjacent communities or high-risk areas, such as the island's mountainous regions (e.g. Ruiz-Mirazo *et al*, 2011). This strategy has already been implemented by the Ministry of Agriculture, Rural Development and the Environment, in collaboration with the Department of Forests and the Department of Agriculture, using goats since 2025 in the Pyrgos Tyllirias forestry area. Promoting the breed's meat products in the local market could also create niche economic opportunities and provide further incentives for conservation. Since the production efficiency of the indigenous cattle breed of Cyprus is relatively low, added value should be created by linking breed preservation to cultural and eco-tourism initiatives, such as farm visits (Pastrana *et al*, 2020). These approaches can improve profitability and, in turn, support long-term conservation efforts.

There are multiple compelling reasons to conserve the indigenous cattle breed of Cyprus, leaving little room for further debate. Situated in a recognized climate change hotspot (Zittis *et al*, 2020), Cyprus must prioritize the preservation of its valuable AnGR, which hold the potential to adapt to increasingly harsh environmental conditions, just as prehistoric farmers successfully managed livestock thousands of years ago. We hope that ongoing advances in palaeogenomics will enhance the recovery and analysis of ancient DNA, not only from cattle but also from other livestock species such as sheep, goats and equids, all of which have played significant roles in the island's prehistoric and more recent agricultural history. Future projects will follow a similar approach, focusing on other indigenous and currently threatened breeds, including the Cyprus fat-tailed sheep, the Machaeras goat and the Cyprus donkey. Equally essential is the collaboration and coordination of such work with partners in nearby regions (see Ligda and Casabianca, 2013). Given the long-standing cultural ties and historical translocation of plants and animals across the Mediterranean as well as the common threats that the region has always been facing, cross-regional collaboration can support the development of large projects, common policies and governance strategies within the Mediterranean Basin. These kinds of platforms that encourage cooperation across the region already exist, including the European Regional Focal Point for Animal Genetic Resources (Martyniuk *et al*, 2021) and the EAAP Mediterranean Working Group (Ligda *et al*, 2022).

Conclusion

The conservation of the indigenous cattle breed of Cyprus, like the broader effort to preserve AnGR, is a complex, multifaceted societal challenge that demands multidisciplinary approaches that integrate scientific methods along with anthropological concepts. Central to this effort is the formation of an engaged community that values animals not only for their productive traits but also for the ecosystem services they provide, their deep-rooted connections to local environments, and their role in enduring cultural traditions. Achieving this vision requires coordinated action from a wide range of stakeholders, including researchers, breeders, public authorities, policymakers, educators, anthropologists, artists and activists. Participatory methods are essential, particularly those that recognize breeders as local experts and incorporate their knowledge, experience and practices. Moreover, public engagement should be an essential component of every project that focuses on the conservation of AnGR. In this context, documentaries can be powerful tools for engaging the wider public in conservation efforts, providing opportunities to safeguard oral histories and cultural traditions and create new spaces for conservation. We hope this integrated approach will be extended to include other indigenous animal breeds in Cyprus, the Eastern Mediterranean and beyond. Strong collaboration across the Mediterranean Basin remains essential to this collective effort.

Author contributions

A. Spyrou and G. Hadjipavlou conceived and designed the project and wrote the first draft of the manuscript. A. Spyrou collected zooarchaeological, archival, and ethnographic data. G. Hadjipavlou and A. C. Dimitriou collected and analyzed

genomic data. A. Spyrou and V. Mattiangeli processed archaeological samples, and D. Diquelou prepared the PCA. V. Mullin analyzed palaeogenomic data and provided overall oversight of the analyses. D.G. Bradley facilitated and oversaw ancient DNA analysis. All authors commented on previous versions of the manuscript.

Acknowledgements

The authors are grateful to the A. G. Leventis Foundation for its generous financial support of this research. We would also like to thank the Cyprus Department of Antiquities for granting permission to export archaeological cattle bones to Ireland for ancient DNA analysis. The ancient DNA work was made possible through Prof. Bradley's ERC Advanced Grant 'AncestralWeave' (Grant Agreement ID: 885729). This work would not have been possible without the support of the Cyprus Department of Agriculture, particularly officers Chryso Demetriou and Charalampos Petsides. Special thanks are due to the Cyprus Press and Information Office for providing access to historical journals and photographs, and to the Cyprus State Archives for granting access to rare archival material. We would also like to express our sincere gratitude to Prof. Ivica Medugorac, Head of the Population Genomics Group at the Ludwig Maximilian University of Munich, for generously providing us with samples from Turkey and Greece. We are deeply indebted to the Cyprus Local Cow Breed's Society for facilitating access to the animals and for generously sharing their stories with us. Finally, our sincere thanks go to Dr George Papageorgiou, veterinarian, whose support and guidance opened many doors throughout the course of this project.

Conflict of interest statement

The authors declare no conflict of interest

Ethics statement

All participating farm owners consented to sample animals for experimentation purposes. Nasal swab samples were collected to minimize stress and avoid harm to the animals. All experimental protocols were approved by the Agriculture Research Institute of Cyprus, which is part of the Ministry of Agriculture, Rural Development and Environment. All methods were carried out in accordance with relevant guidelines and regulations. All methods are reported in accordance with ARRIVE guidelines.

References

Achilli, A., Bonfiglio, S., Olivieri, A., Malusà, A., Pala, M., Kashani, B. H., Perego, U. A., Ajmone-Marsan, P., Liotta, L., Semino, O., Bandelt, H.-J., Ferretti, L. and Torroni, A. (2009). The multifaceted origin of taurine cattle reflected by the mitochondrial genome. *PLoS One* 4, e5753. doi: https://doi.org/10.1371/journal.pone.0005753

Apostolou, D. (2018) Traditional Methods of Cereal Cultivation in Cyprus. *Agrotis* 475, 52-60.

Averett, E. W. (2020). Blurred boundaries: Zoomorphic masking rituals and the human-animal relationship in ancient Cyprus. *World Archaeology* 52, 724–745. doi: https://doi.org/10.1080/00438243.2021.1900903

Boessenkool, S., Hanghøj, K., Nistelberger, H.M., Der Sarkissian, C., Gondek, A.T., Orlando, L., Barrett, J.H. and Star, B. (2017). Combining bleach and mild predigestion improves ancient DNA recovery from bones. *Molecular Ecology Resources* 17, 742–751. doi: https://doi.org/10.1111/1755-0998.12623

Boyazoglu, J., Hatziminaoglou, Y. (2002). Livestock genetic resources and production systems: A Mediterranean overview. *Archivos Latinoamericanos de Producción Animal*, 10(1), 62–67.

Constantinides, P. 1955. Production of Clean Milk. *Agrotis* 9(4), 16-17.

Croft, P. (2006). Animal Bones. In Marki Alonia: An Early and Middle Bronze Age Settlement in Cyprus: Excavations 1995–2000, eds. D. Frankel and JM Webb (Paul Åströms förlag), 263–281.

Dabney, J., Meyer, M. (2019). Extraction of highly degraded DNA from ancient bones and teeth. In Ancient DNA: methods and protocols, eds B. Shapiro, A. Barlow, P.D. Heintzman, M. Hofreiter, J.L.A. Paijmans, A.E.R. Soares (New York, NY: Springer New York), 25-29.

Davis, S.J.M. (2003). The zooarchaeology of Khirokitia (Neolithic Cyprus), including a view from the mainland. In Le Néolithique de Chypre, Bulletin de Correspondence Hellénique, Supplément 43, eds J. Guilaine, A. Le Brun (École Française d'Athènes, Athens), 253–268.

da Gama, L. T. (2006). Animal genetic resources and sustainable development in the Mediterranean region. In Animal products from the Mediterranean area, eds. J.M.C. Ramalho Ribeiro, A.E.M. Horta, C. Mosconi and A. Rosati (Wageningen: Wageningen Academic), 125–135.

de Mas Latrie, L., 1855. History of the island of Cyprus under the reign of the princes of the house of Lusignan (Vol. 3). Imperial Printing Office.

Department of Agriculture (DAG) 2024. Yearly review of the Cyprus Dairy Cow Sector and the Local Cattle Breed. ISSN 2980-5325, Ministry of Agriculture, Rural Development and Environment.

Dimitriou, A.C., Antoniou, A., Alexiou, I., Poulakakis, N., Parmakelis, A. and Sfenthourakis, S. (2022). Diversification within an oceanic Mediterranean island: Insights from a terrestrial isopod. *Molecular Phylogenetics and Evolution*, 175, p.107585. doi: https://doi.org/10.1016/j.ympev.2022.107585

Dimitriou, A.C., Maimaris, G. and Hadjipavlou, G. (2024b). Assessment of breeding nuclei contributions to the genetic diversity and population structure of the Cyprus Chios sheep. *Scientific Reports*, 14(1),1-12. doi: https://doi.org/10.1038/s41598-024-81678-3

Dimitriou, C.A., Spyrou, A. & Hadjipavlou, G. (2024a). First insights in the genetic structure of Cyprus's indigenous cattle. Book of Abstracts for the 75th Annual Meeting of the European Federation of Animal Science.

Felius, M., Beerling, M.L., Buchanan, D.S., Theunissen, B., Koolmees, P.A. and Lenstra, J.A., 2014. On the history of cattle genetic resources. *Diversity*, 6(4),705-750. doi: https://doi.org/10.3390/d6040705

Flori, L., Moazami-Goudarzi, K., Alary, V., Araba, A., Boujenane, I., Boushaba, N., Casabianca, F., Casu, S., Ciampolini, R., Coeur D'Acier, A. and Coquelle, C. (2019). A genomic map of climate adaptation in Mediterranean cattle breeds. *Molecular ecology*, 28(5),1009-1029. doi: https://doi.org/10.1111/mec.15004

Fraisl, D., Hager, G., Bedessem, B., Gold, M., Hsing, P.Y., Danielsen, F., Hitchcock, C.B., Hulbert, J.M., Piera, J., Spiers, H. and Thiel, M., 2022. Citizen science in environmental and ecological sciences. *Nature reviews*

methods primers, 2(1), 64. doi: https://doi.org/10.1038/s43586-022-00144-4

Gamba, C., Jones, E.R., Teasdale, M.D., McLaughlin, R.L., Gonzalez-Fortes, G., Mattiangeli, V., Domboróczki, L., Kővári, I., Pap, I., Anders, A. and Whittle, A. (2014). Genome flux and stasis in a five millennium transect of European prehistory. *Nature communications*, 5(1), 5257. doi: https://doi.org/10.1038/ncomms6257

Gandini, G.C., Villa, E. (2003). Analysis of the cultural value of local livestock breeds: a methodology. *Journal of Animal Breeding and Genetics* 120 (1), 1-11. https://doi.org/10.1046/j.1439-0388.2003.00365.x

Gandini, G., Avon, L., Bohte-Wilhelmus, D., Bay, E., Colinet, F.G., Choroszy, Z., Díaz, C., Duclos, D., Fernández, J., Gengler, N. and Hoving-Bolink, R. (2010). Motives and values in farming local cattle breeds in Europe: a survey on 15 breeds. *Animal Genetic Resources/Resources génétiques animales/Recursos genéticos animales*, 47, 45-58. doi: https://doi.org/10.1017/S2078633610000901

Green, R.E., Krause, J., Briggs, A.W., Maricic, T., Stenzel, U., Kircher, M., Patterson, N., Li, H., Zhai, W., Fritz, M.H.Y. and Hansen, N.F. (2010). A draft sequence of the Neandertal genome. *Science*, 328(5979), pp.710-722. doi: https://doi.org/10.1126/science.1188021

Hall, S.J. (2019). Livestock biodiversity as interface between people, landscapes and nature. *People Nat*, 1(3), 284-290. doi: https://doi.org/10.1002/pan3.23

Hadjipavlou, G., Ligda, C. (2022). Diversity of animal genetic resources for resilient farming systems. In Hadjipavlou, G., and Ligda, C. (eds) Addressing the challenges of agro-pastoral farming systems to strengthen their resilience. Options Méditerranéennes: Série A. Séminaires Méditerranéens 129, 11–15. Zaragoza: CIHEAM. Proceedings of the 71st Annual Meeting of the European Federation of Animal Science (EAAP). URL: http://om.ciheam.org/om/pdf/a129/00008091.pdf

Haudricourt A. G. (1962) Domestication des animaux, culture des plantes et traitement d'autrui, L'Homme, 2(1), 40-50.

Hoffmann, I. (2010). Climate change and the characterization, breeding and conservation of animal genetic resources. *Animal Genetics*, 41, 32-46. doi: https://doi.org/10.1111/j.1365-2052.2010.02043.

Holmes, M., Hamerow, H. and Thomas, R., 2021. Close companions? A zooarchaeological study of the human–cattle relationship in medieval England. *Animals*, 11(4), 1174.

Horwitz, L.K, Tchernov,E., and Hongo, H. 2004. The domestic status of the early Neolithic fauna of Cyprus: a view from the mainland. In Neolithic Revolution: New Perspective on Southwest Asia in light of Recent Discoveries on Cyprus, ed E. Peltenburg and A. Wasse (Levant Supplementary Series I. Oxbow Books, Oxford) 35–48.

Kaniewski, D., Marriner, N., Bretschneider, J., Jans, G., Morhange, C., Cheddadi, R., Otto, T., Luce, F. and Van Campo, E. (2019). 300-year drought frames Late Bronze Age to Early Iron Age transition in the Near East: new palaeoecological data from Cyprus and Syria. *Regional Environmental Change*, 19, 2287-2297.

Karageorghis V. (2005). Excavations at Kition VI. The Phoenician and later levels, part I. Nicosia.

Klerides, N. (2017). Istories ton Zoon tis Kyprou. (Lefkosia : Politistiko Idryma Nearchou Kleride).

Köhler-Rollefson, I. (1997). Indigenous practices of animal genetic resource management and their relevance for the conservation of domestic animal diversity in developing countries. *Journal of Animal Breeding and Genetics*, 114(1-6), 231-238.

Köhler-Rollefson, I. 2001. Intellectual property rights regime necessary for traditional livestock raisers. Indigenous Knowledge and Development Monitor, 9(1), 12-15.

Köhler-Rollefson, I. (2003). Community-based management of animal genetic resources–with special reference to pastoralists. In Proceedings of the workshop held in Mbabane, Swaziland (Rome).

Ligda C., Casabianca F. (2013). Adding value to local breeds: challenges, strategies and key factors. *Animal Genetic Resources* 53, 107–116. doi: https://doi.org/10.1017/S207863361300026X

Ligda C., Tsiokos D., Hadjipavlou G., Djemali M., El Fadili M., Charvolin-Lemaire E., Sturaro E. (2022). Documenting animal genetic resources in the Mediterranean: interaction and cooperation in the region. In Addressing the challenges of agro-pastoral farming systems to strengthen their resilience, eds G. Hadjipavlou, C. Ligda (Options Méditerranéennes: Série A. Séminaires Méditerranéens; n. 129) Zaragoza, 49-56p.

Martyniuk, E., Matlova, V., Sæther, N., Bojkovski, D., Puzin, L., Polak, G., Hiemstra, S.J., Berger, B., Bormann, J., Stålhammar, E.M. and Stojanović, S. (2021). PersPectives of NatioNal coordiNators for aNimal GeNetic resources oN the euroPeaN reGioNal focal PoiNt: survey results. *Annals of Animal Science*, 21(3), pp.853-868. doi: https://doi.org/10.2478/aoas-2020-0120

Mendelsohn, R. (2003). The challenge of conserving indigenous domesticated animals. *Ecological Economics*, 45(3), 501-510. doi: https://doi.org/10.1016/S0921-8009(03)00100-9

Meyer, M., Kircher, M. (2010). Illumina sequencing library preparation for highly multiplexed target capture and sequencing. Cold Spring Harb. Protoc., db.prot 5448.

Neophytou, A. (2014). Laiki latreia stin paradosiaki koinonia tis Kyprou. Oi giortes tou dodekaimerou (Publications of the Cyprus Research Center-Nicosia, Cyprus).

Ohnefalsch-Richter, M.H. (1913). Griechische Sitten und Gebräuche auf Cypern: mit Berücksichtigung von Naturkunde und Volkswirtschaft sowie der Forschritte unter englischer Herrschaft, dargestellt. D. Reimer (E. Vohsen).

Ovaska, U. (2017). Genes, gastronomy and gratitude: The development and future of the conservation of native breeds. Unpublished PhD Dissertation.

Ovaska, U., Soini, K., 2017. Local breeds–Rural heritage or new market opportunities? Colliding views on the conservation and sustainable use of landraces. *Sociologia Ruralis*, 57, 709-729. doi: https://doi.org/10.1111/soru.12140

Ovaska, U., Bläuer, A., Kroløkke, C., Kjetså, M., Kantanen, J. and Honkatukia, M. (2021). The Conservation of Native Domestic Animal Breeds in Nordic Countries: From Genetic Resources to Cultural Heritage and Good Governance. *Animals* 2021, 11, 2730. doi: https://doi.org/10.3390/ani11092730.

Panaretos, A. 1967. Kypriaki Georgiki Laografia. Publications of the Cooperative Central Bank of Cyprus Ltd, Nicosia.

Papachristou, D.A., 2023. Genetic identification of Greek bovine breeds/populations using molecular markers. Ph.D. thesis, University of Athens, Athens.

Papachristou, D., Koutsouli, P., Laliotis, G.P., Kunz, E., Upadhyay, M., Seichter, D., Russ, I., Gjoko, B., Kostaras, N., Bizelis, I. and Medugorac, I. (2020). Genomic diversity and population structure of the indigenous Greek and Cypriot cattle populations. *Genetics Selection Evolution*, 52, 1-23. doi: https://doi.org/10.1186/s12711-020-00560-8

Pastrana, I. C., Navas González, F.J., Ciani, E., Nogales Baena, S. and Delgado Bermejo, J.V. (2020). Camel genetic resources conservation through tourism: a key sociocultural approach of camelback leisure riding. *Animals*, 10(9), 1703. doi: https://doi.org/10.3390/ani10091703

Pillar, V.D., Overbeck, G.E. (2025). Grazing can reduce wildfire risk amid climate change. *Science*, 387(6730), eadu7471. doi: https://doi.org/10.1126/science.adu7471

Poulakakis, N., Kapli, P., Kardamaki, A., Skourtanioti, E., Göçmen, B., Ilgaz, Ç., Kumlutaş, Y., Avci, A., Lymberakis, P. (2013). Comparative phylogeography of six herpetofauna species in Cyprus: late Miocene to Pleistocene colonization routes. *Biological Journal of the Linnean Society* 108, 619–635. doi: https://doi.org/10.1111/j.1095-8312.2012.02039.x

Pritchard, K. J., Stephens M. and Donnelly, P. (2000). Inference of Population Structure Using Multilocus Genotype Data *Genetics*, 155 (2), 945–959, https://doi.org/10.1093/genetics/155.2.945

Redding, R.W. (2024). A view from the herd: cattle, sheep, goats, and pigs in pharaonic Egypt: a primer for Egyptologists and archaeologists. Lockwood Press, 210p.

Rischkowsky, B., Pilling, D. (2007). The state of the world's animal genetic resources for food and agriculture. Food and Agriculture Organisation (Rome: Italy). url: https://www.fao.org/3/a1250e/a1250e.pdf

Rizopoulou-Egoumenidou, E. (2009). Tanning in Cyprus from the 16th to the 20th century: from traditional tanneries to modern industries. (Nicosia: Kentron Epistēmonikōn Ereunōn).

Ruiz-Mirazo, J., Robles, A.B. and González-Rebollar, J.L. (2011). Two-year evaluation of fuelbreaks grazed by livestock in the wildfire prevention program in Andalusia (Spain). *Agriculture, Ecosystems & Environment*, 141(1-2), 13-22. doi: https://doi.org/10.1016/j.agee.2011.02.002

Sabiescu, A. G. (2020). Living archives and the social transmission of memory. Curator: *The Museum Journal* 63, 497–510. doi: https://doi.org/10.1111/cura.

Savile, A.R. 1878. Cyprus. Compiled in the Intelligence Branch, Quarter-Master-General's Department, Horse Guards. H.M. Stationery Office, London. 171 pp. url: https://sylviaioannoufoundation.org/en/collection/digital-library/b2181/

Scherf, B. D. (2000). World watch list for domestic animal diversity (3rd ed.). Food and Agriculture Organization of the United Nations. https://www.fao.org/4/x8750e/x8750e.pdf

Scott-Stevenson, E., 1880. Our home in Cyprus. (London: Chapman and Hall).

Soini, K., Ovaska, U. and Kantanen, J. (2012). Spaces of conservation of local breeds: the case of Yakutian cattle. *Sociologia Ruralis*, 52(2), pp.170-191. doi: https://doi.org/10.1111/j.1467-9523.2012.00551.x

Spyrou A. (2021). From south Asia to the eastern Mediterranean: the appearance of zebu cattle in the iconography of Bronze Age Cyprus and some consideration for future research. In Archaeology in the smallest realm: micro analyses and methods for the reconstruction of Cyprus early societies, ed M. Amadio (Artemide: Rome), 154-166.

Spyrou, A., Roberts, P., Bleasdale, M., Lucas, M., Crewe, L., Simmons, A. and Webb, J., 2024. Human-cattle interactions in PPNB and Early/Middle Bronze Age Cyprus: integrating zooarchaeological and stable isotope data. Archaeological and Anthropological Sciences, 16(5), 64. doi: https://doi.org/10.1007/s12520-024-01969-8

Surridge, B. J. (1930). A survey of rural life in Cyprus. (Nicosia: Government Printing Office).

Thiringer, M.M. (2020). An Egyptian's Best Friend? An Analysis and Discussion of the Depiction of the Domestic Dog in Ancient Egypt. Ph.D. thesis, University of Memphis, Memphis, USA.

Thomas, R. and Johannsen, N., 2011. Articular depressions in domestic cattle phalanges and their archaeological relevance. *International Journal of Paleopathology*, 1(1), pp.43-54.

Utsunomiya, Y.T., Milanesi, M., Fortes, M.R.S., Porto-Neto, L.R., Utsunomiya, A.T.H., Silva, M.V.G.B., Garcia, J.F. and Ajmone-Marsan, P. (2019). Genomic clues of the evolutionary history of *Bos indicus* cattle. *Animal genetics,* 50 (6), 557-568. doi: https://doi.org/10.1111/age.12836

Van Dijk, R.M. (2013). The use of bucrania in the architecture of First Dynasty Egypt. *Journal for Semitics*, 22(2), 449-463. doi: https://hdl.handle.net/10520/EJC147422

Vayanou, P., Papatheodoulou, A., Hellicar, M., Hadjistylli, M., Perdiou, A., Charilaou, P. and Sergides, L. (2024). Sustainability plan for Cape Pyla and Akrotiri Peninsula: Recommendations & measures. DPLUS141: Habitat restoration and wise use for Akrotiri and Cape Pyla. BirdLife Cyprus; Terra Cypria; Sovereign Base Areas Administration Environment Department (SBBA ED); Royal Society for the Protection of Birds (RSPB).

Verdugo, M.P., Mullin, V.E., Scheu, A., Mattiangeli, V., Daly, K.G., Maisano Delser, P., Hare, A.J., Burger, J., Collins, M.J., Kehati, R. and Hesse, P. (2019). Ancient cattle genomics, origins, and rapid turnover in the Fertile Crescent. *Science*, 365(6449), pp.173-176. doi: https://doi.org/10.1126/science.aav1002

Vigne, J.D., Zazzo, A., Cucchi, T., Carrère, I., Briois, F. and Guilaine, J. (2014). The transportation of mammals to Cyprus sheds light on early voyaging and boats in the Mediterranean Sea. *Eurasian Prehistory*, 10(1-2), 157-176.

Xioutas, P. (1978). KypriakI Laographia ton Zoôn, Vol. 6 (Nicosia : Kentron Episfëmonikôn Meletôn).

Zittis, G., Bruggeman, A. and Camera, C., 2020. 21st century projections of extreme precipitation indicators for Cyprus. *Atmosphere*, 11(4), p.343. doi: https://doi.org/10.3390/atmos11040343

Genetic Resources

ORIGINAL ARTICLE

Genetic Resources (2026), 7 (13), 129–142
DOI: 10.46265/genresj.ASMH5957
https://www.genresj.org
ISSN: 2708-3764

Expanding the genetic diversity of chickpeas from the Ukrainian genebank to new agricultural systems

Nadiia Vus[a,b,*], Olha Bezuhla[b], Serhii Sylenko[c], Antonina Vasylenko[b], Viacheslav Sichkar[d], Mykola Kondratenko[e], Margarita Barylko[f]

[a] *Agroécologie, INRAE, Institut Agro, Univ. Bourgogne, Univ. Bourgogne Franche-Comté, F-21000 Dijon, France*
[b] *Yuriev Plant Production Institute of NAAS, Kharkiv, Ukraine*
[c] *Ustymivka Experimental Plant Growing Station of Yuriev Plant Production Institute of NAAS, Ustymivka, Hlobynskyi District, Poltava Region, Ukraine*
[d] *Plant Breeding and Genetics Institute – National Center of Seed and Cultivar Investigation, Odesa, Ukraine*
[e] *Institute of Feed Research and Agriculture of Podillya NAAS, Vinnytsia, Ukraine*
[f] *Poltava State Agricultural Research Station named after M.I. Vavilov, Institute of Pig Breeding and Agro-Industrial Production of NAAS of Ukraine, Poltava, Ukraine*

Abstract: Enhancing crop resilience to abiotic factors in an increasingly unstable climate is a critical challenge, one that can be met by actively using the existing gene pool of the crops. A total of 26 new chickpea lines with diverse ancestry were selected from the Ukrainian genebank with the objective of widening the genetic base and creating varieties that are more adaptable to the prevailing climatic conditions in Ukraine. The research was conducted between 2019 and 2021 at four distinct locations: dry forest-steppe, Kharkiv region (Elitne); wet forest-steppe, Poltava region (Ustymivka) and hyper-humid Polissya, Vinnytsia region (Bokhonyky); and the extremely arid conditions of the Odesa region (Khlibodarske, 2020). A GGE-biplot analysis was utilized to assess the adaptability of the chickpea lines to diverse environmental conditions. New breeding lines with an elevated degree of adaptability to the Ukrainian forest-steppe zone were identified. Furthermore, three new sources with potential resistance to hyper-humidity were identified (genotypes 2072, 2067 and 2065). Genotypes 2068 and 2088 have been registered in the NCPGRU as sources of complex valuable traits, and genotype 2087 has been registered as a source of high adaptability. New and existing chickpea accessions from the genebank of Ukraine can serve as the basis for the development of new breeding materials, thus helping solve modern challenges facing breeders.

Keywords: *Cicer arietinum*, adaptation, genotype–environment interaction, resistance, abiotic stress.

Citation: Vus, N., Bezuhla, O., Sylenko, S., Vasylenko, A., Sichkar, V., Kondratenko, M. and Barylko, M. (2026) “Expanding the genetic diversity of chickpeas from the Ukrainian genebank to new agricultural systems”, *Genetic Resources*, 7(13), pp. 129–142. doi: 10.46265/genresj.ASMH5957

Introduction

Climate change is a key factor driving the emergence of biotic and abiotic stresses that adversely affect agricultural development. Mitigating the consequences of this impact is important for the further development of agricultural production (Raza *et al*, 2019; Pixley *et al*, 2023). These measures include the introduction of new crop varieties resilient to changing climatic conditions, adjustments to the geographical distribution of crop cultivation, improvements to crop rotation schemes, and the development of diversification strategies (Hristov, 2020). Although the relocation of agricultural production to areas with more favourable climatic conditions is considered a potential adaptation strategy (Sloat *et al*, 2020), its successful implementation requires the availability of well-adapted crop varieties. Therefore, the exploration and utilization of genetic resources represent a critical component of climate-resilient agriculture, with crop adaptability and diversification identified as major priorities (Mohammadi *et al*, 2023). The inclusion of legumes in crop rotations reduces dependence on mineral nitrogen fertilizers due to their symbiotic relationship with nitrogen-fixing

* Corresponding authors: Nadiia Vus (vus.nadezhda@gmail.com)

Received: 27.11.2025 | Accepted: 16.03.2026 | Published online: 07.04.2026

bacteria, which can increase the productivity of subsequent crops (Lampkin *et al*, 2015). European agriculture allocates only 1.5% of arable land to legume cultivation, compared with the global average of 14.5%. At the same time, the European livestock feed market exhibits a high demand for protein-rich ingredients, most of which are supplied by imports of soybeans and soybean meal, accounting for 87% of total supply (Watson *et al*, 2017). This situation underscores the urgent need to explore alternative legume grains – both domestically produced and imported – to reduce dependence on soybean-based feed. To achieve this, an important strategy is to diversify legume crops beyond those currently widely cultivated and to explore species that are still considered niche crops (Watson *et al*, 2017; Khan *et al*, 2024).

Chickpea (*Cicer arietinum* L.) represents such a crop for Europe. It is a valuable source of protein-rich plant products and is characterized by strong drought tolerance and high seed quality (Berrada *et al*, 2007). Therefore, it has potential to be incorporated into more diverse legume-based crop rotation systems. Chickpea is one of the world's most important legumes, ranking third in production area after beans and cowpeas, with 14.8 million ha (17.2%) (FAOSTAT, 2023).

The main world chickpea breeding goals include increasing yield potential, as well as expanding adaptation to various environmental conditions and increasing resistance to biotic and abiotic stresses (Gaur *et al*, 2007). Genotype-by-environment (G × E) interaction poses a major challenge: varieties bred for high yield in their region of origin often perform poorly elsewhere. This limits their expansion into new regions and undermines stability as local climates shift. This problem is also present in Ukraine. Here, the main chickpea breeding centre, the Plant Breeding and Genetics Institute (PBGI) in Odesa, a region characterized by drought and heat, has developed more than 15 chickpea varieties (The Plant Breeding and Genetics Institute, 2024). Developing new lines from adapted germplasm and evaluating them across novel geographic regions is a key strategy for advancing crop diversification (von Wettberg *et al*, 2018; Varshney *et al*, 2021). The assessment of G × E interactions and the evaluation of the adaptability of source material constitute key stages in the breeding process. The region where the genebank is located (Elitne, Kharkiv region) represents a high-risk area for chickpea cultivation. However, long-term research has shown that certain genetic resource accessions are sufficiently adapted to local conditions. In contrast, the Ustymivka location is atypical for chickpea cultivation, while the Bokhonyky location is excessively waterlogged and therefore potentially unsuitable for this crop.

Newly developed chickpea accessions, selected from those that survived in extreme environmental conditions in previous years, drawn from both existing accessions and hybrid combinations, were evaluated under non-standard conditions. The objective of this study was to evaluate newly developed chickpea accessions under contrasting and atypical environmental conditions in order to identify genotypes suitable for the development of new varieties and for expanding chickpea cultivation into non-traditional growing regions (Rebollo *et al*, 2023; Fritsche-Neto *et al*, 2025).

This study evaluates the adaptive potential of chickpea genetic resources developed in the Ukrainian genebank under contrasting environmental conditions and provides reference data to support future germplasm research and utilization.

Materials and methods

Plant material

The study examined 26 chickpea (*Cicer arietinum* L.) breeding lines developed at the Yuriev Plant Production Institute of the National Academy of Agrarian Sciences of Ukraine (Kharkiv, Ukraine; YPPI NAAS) within two research programmes: resistance to *Ascochyta rabiei* and high-yield selection. The breeding material originated from two sources: (1) elite lines selected over multiple years from the base chickpea collection of the National Center for Plant Genetic Resources of Ukraine (NCPGRU) that survived under Ascochyta epiphytotic pressure; (2) lines derived from targeted hybridization; a third group consists of standard accessions that are well adapted to the local environmental conditions.

All genotypes underwent a multi-stage selection process aimed at identifying lines with resistance to Ascochyta under field conditions. Parental combinations were chosen based on preliminary evaluations of chickpea adaptability to the environmental conditions of the Eastern Forest-Steppe zone of Ukraine (Table 1). Detailed passport data for all genotypes are available via their catalogue numbers at NCPGRU or through the international plant genetic resources for food and agriculture (PGRFA) information platform, Genesys-PGR (Genesys-PGR, 2025).

Field trials

The field experiments were carried out at four locations in Ukraine. Three-year trials (2019–2021) were conducted at Elitne, Ustymivka, and Bokhonyky. An additional site, Khlibodarske, was included in 2020 to evaluate genotype performance under extremely arid climatic conditions (Figure 1).

A summary of the climatic and weather characteristics of each location is presented in Supplemental Table 1.

A 4-year crop rotation was applied, with winter wheat systematically used as the preceding crop for chickpea. The experiments were conducted in accordance with the Methodical Recommendations for Studying the Genetic Resources of Grain Legumes (Kobyzeva *et al*, 2016). Each experimental plot covered an area of 1m², and the sowing scheme was 30cm × 10cm, comprising three rows of 10 plants (30 plants per plot). Seeds were sown manually. Phenological observations, analysis of the crop structure of accessions were carried out according to methodological recommendations for the study of genetic resources of leguminous crops (Kobyzeva *et al*, 2016) and taking into account the chickpea traits ontology (Rani Das *et al*, 2024).

Following the methodology for genetic resource evaluation, and considering the absence of replications, one plot of the specific reference accession selected for the respective location was included every 20 plots to ensure accurate field assessment of all tested entries with normalization of the yield of all tested accessions relative to the average for the block of standard-accessions. To correct for field inhomogeneity, a standard shift correction was used, using a correction through the average of the standard block (Rozanna and Krasnokutskyy 123). Hand harvesting was performed. Before threshing, plant height (M1) and the height of the lowest pod attachment (M2) were measured using a ruler under laboratory conditions. Plants were then threshed individually using a laboratory thresher. Plot yield (P3) was

Table 1. Chickpea genotypes included in the study and their origin. NCPGRU, National Center for Plant Genetic Resources of Ukraine. *, Accession that does not have a catalogue number but only a collection working number.

Catalogue number of NCPGRU	Genotype name	Pedigree
1) Individual selection		
UD0502239	2065	UD0500093 – Local variety, India
UD0502240	2066	UD0500134 – Local variety, Spain
UD0502241	2067	UD0500733 – LR 17 1, Syria
UD0502245	2068	UD0500864 – Flip 99-55c, Syria
UD0502246	2069	UD0502065 – LUH 106-07, Ukraine
UD0502247	2070	UD0502093 – Local variety, Ukraine
UD0502248	2071	UD0500196 – Local variety, Azerbaijan
UD0502249	2072	UD0500240 – ILC 3279, Syria
UD0502242	2073	UD0500444 – Dniprovskyi vysokoroslyi, Ukraine
UD0502250	2074	UD0502112 – Local variety, Ukraine
UD0502251	2075	UD0502111 – Local variety, Russia
UD0502252	2076	UD0501504 – Local variety, Ukraine
UD0502253	2077	UD0500671 – K 4-select, India
2) Breeding lines		
UD0502254	2078	Rozanna (UD0500424) X Krasnokutskyi 123 (UD0500101)
UD0502255	2079	Rozanna (UD0500424) X Krasnokutskyi 123 (UD0500101)
UD0502256	2080	Antei (UD0500735) X Krasnokutskyi 123 (UD0500101)
UD0502257	2081	Antei(UD0500735) X Krasnokutskyi 123 (UD0500101)
UKR001:02082*	2082	Krasnokutskyi 123(UD0500101) X Antei (UD0500735)
UD0502258	2083	Krasnokutskyi 123(UD0500101) X Antei (UD0500735)
UD0502259	2084	Krasnokutskyi 123(UD0500101) X Antei (UD0500735)
UD0502260	2085	Krasnokutskyi 123(UD0500101) X Antei (UD0500735)
UD0502243	2086	Luhanets (UD0500102) X Antei (UD0500735)
UD0502261	2087	Luhanets (UD0500102) X Antei (UD0500735)
UD0502262	2088	Luhanets (UD0500102) X Antei (UD0500735)
UD0502263	2089	Luhanets (UD0500102) X Antei (UD0500735)
UD0502264	2090	Luhanets (UD0500102) X Antei (UD0500735)
3) Standard accessions (adapted to local conditions)		
UD0500424	Rozanna	For Elitne and Bokhonyky
UD0500101	Krasnokutskyi 123	For Elitne and Bokhonyky
UD0502025	Odisei	For Ustymivka
UD0500736	Pamiat	For Khlibodarske

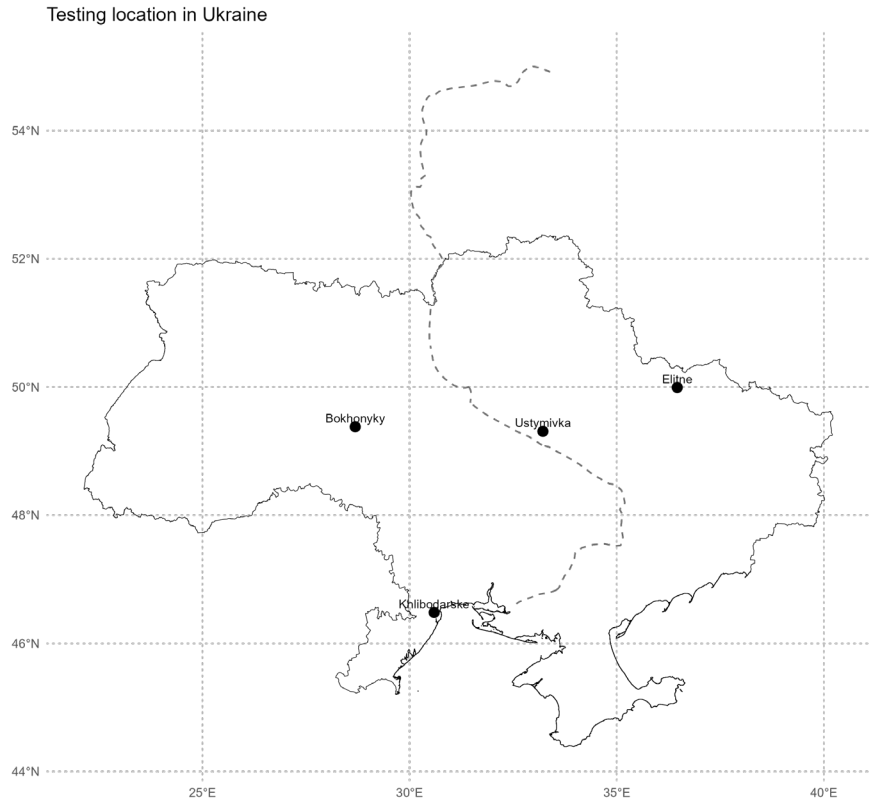


Figure 1. Locations of the study sites in Ukraine. The dashed line indicates the Dnipro River.

calculated based on the seed weight recorded immediately after threshing. Seed yield per plant (P1) was calculated by dividing the yield (P3) by the number of plants harvested. The weight of 100 seeds (P2) was determined by manual counting and weighing in laboratories in five replicates.

Meteorological conditions were recorded directly at the experimental stations during the study period. The duration of phenological phases is noted for each accession: F2, days from germination to 50% flowering (vegetative period); F8, days from 50% flowering to full maturity (generative period); F5, days from germination to full maturity (complete vegetation period). For these periods meteorological variables were assessed for each genotype. The assessment included the cumulative effective temperatures more than 10°C (TF5) and total precipitation (PF5) for the entire growing period (from germination to full maturity). These parameters were also calculated separately for the vegetative phase (from germination to 50% flowering) and the generative phase (from 50% flowering to full maturity) for each accession (TF2 – PF2 and TF8 – PF8, respectively).

Statistical analysis and data visualization

All statistical analyses and data visualization were performed using R software version 4.2.2 (R Core Team, 2023). Data preprocessing and basic statistical calculations were conducted using the 'openxlsx' (Schauberger & Walker, 2022), 'tidyverse', and 'rlang' packages (Henry & Wickham, 2023). Graphical visualization was carried out using 'ggplot2' (Wickham, 2016), with figure composition supported by 'patchwork' (Pedersen, 2024) and label optimization by 'ggrepel' (Slowikowski, 2024). Spatial visualization and mapping of experimental sites were performed using the 'terra' package (Hijmans, 2025).

Multiple factor analysis (MFA) was applied to explore relationships among groups of quantitative and qualitative variables using the 'FactoMineR' package (version 2.7) implemented through the 'Factoshiny' graphical interface (Vaissie *et al*, 2023). A multivariate analysis was performed to provide a comprehensive assessment of the accessions and environmental conditions. All quantitative traits were grouped into four categories:

1. Yield traits: P1, seed yield per plant; P2, 100-seed weight; P3, seed yield per area
2. Morphological traits: M1, plant height; M2, height of the lowest pod
3. Phenological traits: F2, days from germination to 50% flowering; F5, days from germination to full maturity; F8, days from 50% flowering to full maturity
4. Meteorological variables: TF2, sum of temperatures >10°C during the vegetative period; PF2, sum of precipitation during the vegetative period; TF8, sum of temperatures >10°C during the generative period; PF8, sum of precipitation during the generative period; TF5, sum of temperatures >10°C from germination to full maturity; PF5, sum of precipitation from germination to full maturity.

Three qualitative variables were included as supplementary: genotype, year, and location (place).

Quantitative variable groups were standardized (mean = 0, standard deviation = 1), and the analysis was based on the correlation matrix. Qualitative variables were included

as supplementary variables and projected onto the factor space to facilitate the interpretation of relationships among genotypes, years and environments.

Ecological evaluation of genotypes was performed using the 'metan' package (Olivoto & Lúcio, 2020). Replicated trials could not be conducted due to the limited amount of seeds. However, the package allows the analysis of multi-environment trials even when replications are not available. Genotype adaptability and stability across environments were assessed using GGE biplot analysis with the following settings: Scaling = 0 (no scaling of eigenvalues) and Centering = 2 (environment-centred). Singular value partitioning (SVP) was set to 1 for genotype performance and stability assessment based on environmental variation, and SVP = 3 was applied for the 'which-won-where' pattern analysis (Olivoto, 2023). Genotype ranking was conducted using the GGE biplot ranking view (type 8), where genotypes were compared with a virtual 'ideal genotype' defined by maximum mean performance and absolute stability. Ranking was based on Euclidean distance to this reference point (Yan *et al*, 2007).

Additionally, regression coefficients (b_i) and residual standard deviations were estimated using the Finlay–Wilkinson regression model (Finlay & Wilkinson, 1963). Linear models were fitted using the lm() function from the stats package implemented in the R environment.

Results

Meteorological conditions during the study period showed considerable variability, allowing for a comprehensive evaluation of the chickpea lines (Figure 2).

Bokhonyky, representing an extra-wet environment, exhibited high precipitation (≥ 240mm) and low temperatures at the beginning of the growing season in all three study years. Khlibodarske, identified as an extra-dry location in 2020, experienced extreme drought throughout the growing season, although temperatures were not excessively high. In Elitne and Ustymivka, environmental conditions varied widely: from extremely dry (Elitne in 2019 and 2021; Ustymivka in 2020) to overly wet, approaching those of Bokhonyky (Elitne in 2020; Ustymivka in 2019).

To compare the response of chickpea yield to annual variation, the yield of standard accessions was evaluated at each observation point. In Elitne and Bokhonyky, the standards were Rozanna and Krasnokutskyi 123; in Ustymivka, the standard was Odisei; and in Khlibodarske it was Pamiat, all selected for optimal adaptation to local conditions. The mean yield of standard-accessions in Elitne indicated that all study years were favourable for chickpea, with high productivity (395g/m² in 2019; 295.8g/m² in 2020; 367.9g/m² in 2021). In contrast, weather conditions in Ustymivka were atypical, resulting in consistently lower yields for the standard Odisei (220.1g/m² in 2019; 223.8g/m² in 2020; 222.2g/m² in 2021). Lower yields were also recorded in Bokhonyky (194.1g/m² in 2019; 134g/m² in 2020; 195.5g/m² in 2021). In Khlibodarske, the standard Pamiat produced only 71.3g/m², highlighting the severity of environmental stress for chickpea. Across the three primary test sites, under waterlogged conditions in Bokhonyky, the mean yield of chickpea accessions was significantly lower than in Ustymivka and Elitne (Figure 3).

Analysis of variance (ANOVA) confirmed significant differences for all observed traits between the three main test sites and across study years. Detailed agronomic and phenological data are presented in Table 2.

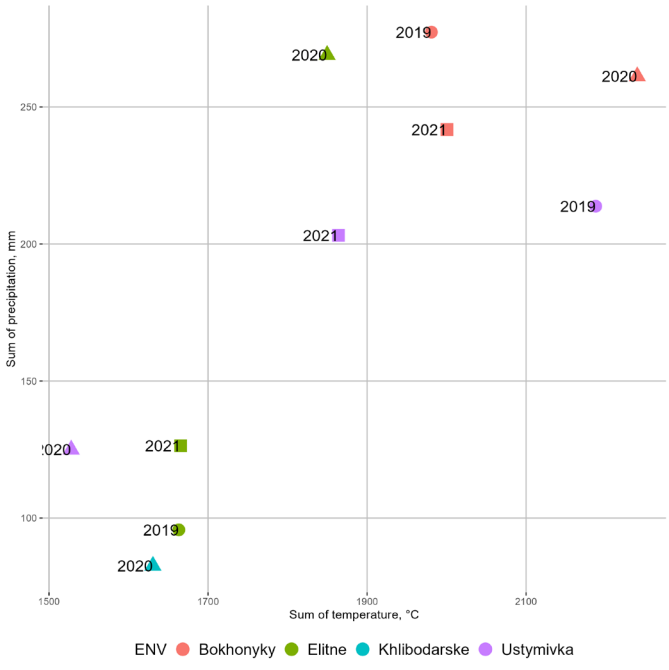


Figure 2. Meteorological conditions during the chickpea growing season.

Figure 3. Yield fluctuations of chickpea accessions at three test sites (2019–2021). Points represent standard-accession means.

Table 2. Analysis of variance for 26 chickpea accessions in ecological testing (2019–2021) at three locations: Elitne (Kharkiv region), Ustymivka (Poltava region) and Bokhonyky (Vinnytsia region). [1], data demonstrated in the format mean (min-max). Statistically significant: *** at P < 0.001; ** at P < 0.01; * at P < 0.05

Location	Elitne			Ustymivka			Bokhonyky			Source of variation	
Year	2019	2020	2021	2019	2020	2021	2019	2020	2021	Location	Year
Seed yield per plant (P1), g	13.12 (7.60–21.20)[1]	10.90 (7.61–16.68)	17.96 (13.80–25.51)	21.88 (8.6–33.5)	8.66 (1.6–28.8)	13.25 (3–23.5)	3.82 (0–14.08)	9.80 (0–30.32)	6.96 (2–12.25)	66.86***	11.04***
Weight of 100 seeds (P2), g	31.69 (23.69–41.82)	31.0 (23.0–42.4)	31.43 (22.19–41.58)	29.59 (22.0–47.11)	28.89 (21.0–39.0)	33.58 (26.0–49.0)	18.99 (0–5.12)	25.49 (0–37.9)	26.09 (20.7–31.0)	29.925***	4.568*
Seed yield per area (P3), g/m2	447.92 (240.8–591.6)	343.96 (194.07–551.6)	461.80 (331.23–586.95)	390.92 (102–615)	412.12 (170–780)	501.76 (220–790)	80.47 (0–300.04)	91.44 (0–239.89)	119.94 (18.92–196.37)	204.99***	9.596***
Plant height (M1), cm	40.35 (27.2–54)	41.35 (26–56.8)	57.42 (40.8–73.8)	48.58 (40–65)	55.81 (49.4–65)	50.70 (36–65)	41.45 (0–80)	72.38 (0–105)	61.25 (45–85)	7.957***	19.88***
Height of the lowest pod (M2), cm	24.3 (12.4–36.4)	20.01 (11.4–28.8)	35.96 (24.2–46.6)	24.49 (17.8–40)	28.39 (22.6–35)	26.22 (16.5–37.5)	25.31 (0–49)	37.86 (0–55)	32.50 (20–45)	5.938**	12.24***
Days from germination to 50% flowering (F2)	34.42 (26–43)	39.69 (36–42)	40.12 (36–46)	38.67 (32–46)	34.25 (32–35)	38 (38–38)	46.2 (43–49)	48.81 (46–51)	50.15 (43–56)	384.73***	25.58***
Days from germination to full maturity (F5)	77.42 (67–86)	91.68 (82–105)	76.5 (72–80)	92.63 (92–99)	72 (72–72)	85.46 (84–89)	106.27 (103–109)	126.95 (124–128)	107.4 (103–113)	1678.90***	67.31***
Days from 50% flowering to full maturity (F8)	43 (31–47)	51.99 (44–65)	36.38 (32–40)	53.96 (46–60)	37.75 (37–40)	47.46 (46–51)	60.07 (56–63)	78.14 (74–81)	57.25 (52–66)	781.7***	108.6***
Sum of temp. >10°C, during the vegetative period, °C (TF2)	659.65 (548.1–833.1)	635.69 (541.7–689.9)	756.02 (655.5–913.5)	923.88 (763.7–1108.5)	613.95 (554.4–633.8)	665.60 (665.6–665.6)	785.37 (733.5–840.4)	625.67 (569.6–673.9)	752.61 (585.5–885.4)	15.91***	131.3***
Sum of prec. during the vegetative period, mm (PF2)	62.74 (34.3–76.8)	137.9 (137.9–137.9)	112.05 (111.7–118.8)	172.87 (133.3–194.9)	83.53 (83.3–83.6)	135.1 (135.1–135.1)	192.61 (182.1–200.5)	163.77 (133.6–166.8)	147.08 (133–154.1)	974.68***	40.49***
Sum of temp. >10°C, during the period from germination to full maturity, °C (TF5)	1662.62 (1443.8–1829.9)	1837.8 (1615.8–2133.4)	1664.74 (1537.1–1777.5)	2187.58 (2175.3–2311.7)	1527.9 (1527.9–1527.9)	1864.07 (1824.2–1960.9)	1980.97 (1929–2023.3)	2239.82 (2190.5–2258.6)	2000.49 (1908.4–2105.8)	316.7***	22.5***
Sum of prec. during the period from germination to full maturity, mm (PF5)	96.27 (75.4–105.2)	269 (269–269)	126.4 (126.4–126.4)	213.8 (208.1–254.9)	125 (125–125)	203.18 (200.9–208.7)	277.32 (268.7–280)	261.22 (230–263.2)	241.78 (234.5–246.5)	4242.7***	419.1***
Sum of temp. >10°C during the generative period, °C (TF8)	1027.35 (861.6–1115.7)	1226.9 (1040.1–1514.5)	934.05 (812.2–1052.6)	1288.48 (1093.9–1438.1)	939.13 (918.5–1001)	1220.07 (1180.2–1316.9)	1195.6 (1099.8–1256.2)	1614.15 (1530.9–1671.6)	1247.88 (1127.3–1438.6)	219.71***	44.11***
Sum of prec. during the generative period, mm (PF8)	34.02 (17.4–41.1)	131.1 (131.1–131.1)	14.63 (12.8–14.7)	40.93 (13.2–74.8)	41.7 (41.7–41.7)	68.08 (65.8–73.6)	84.71 (79.4–86.7)	97.45 (96.4–109.6)	94.7 (86.6–110.4)	473.8***	474.3***
Temperature range	21 (5,9–30,1)	20 (8,9–29,5)	21,5 (9,8–36)	21,4 (9,2–28,4)	20,9 (10,2–33)	20,5 (10,1–30,1)	18,7 (6,3–25,3)	18,2 (6,4–27,6)	18,8 (8,6–26,8)	-	-

Multiple factor analysis (MFA) explained 67.71% of the total variation by the first two dimensions (44.51% and 23.2%, respectively) (Figure 4).

The individuals' plot (Figure 4A) was constructed by projecting individuals onto the first two factorial axes of the MFA, where the coordinates of each individual are defined as a weighted linear combination of the normalized values of the variables on the corresponding axes. The closer the genotypes are, the more similar they are across all groups of variables. The position of an individual relative to the axes reflects its profile on the variables that are most correlated with the corresponding factors. Additional categorical variables, location and year, are also superimposed on the individual plot (for illustrative purposes). Thus, the accessions in the left part of the graph have higher than average yield and its components, which better realized their potential in the conditions of Ustymivka and Elitne. The samples in Bokhonyky are characterized by an extension of the growing season due to waterlogging, but low yield. The MFA revealed a marked distinction between Bokhonyky and the other two sites (Figure 4A). The first axis was driven primarily by weather variables and associated phenophases durations, which differed strongly across locations. The second axis reflected variability in morphological traits. Yield traits showed strong negative correlations with phenophases duration and meteorological conditions (Table 3).

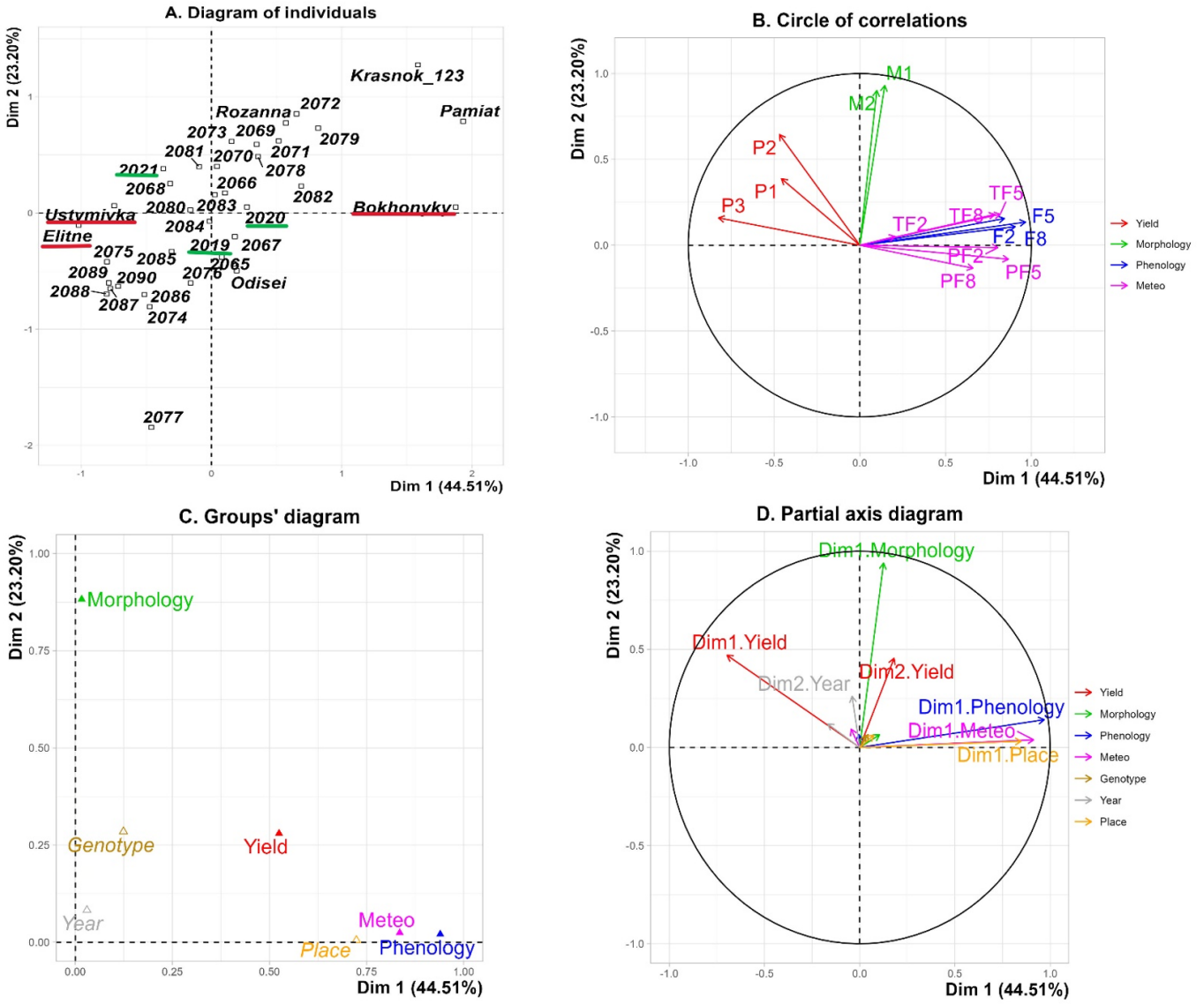


Figure 4. Multiple factor analysis (MFA). In the individual factor map (A. Diagram of individuals), red labels indicate test locations, green labels indicate years, and unlabelled numbers represent genotypes. Colours in other diagrams correspond to trait groups. P1, seed yield per plant; P2, 100-seed weight; P3, seed yield per area, M1, plant height; M2, height of the lowest pod; F2, days from germination to 50% flowering; F5, days from germination to full maturity; F8, days from 50% flowering to full maturity, TF2, sum of temperatures > 10°C during the vegetative period; PF2, sum of precipitation during the vegetative period; TF8, sum of temperatures > 10°C during the generative period; PF8, sum of precipitation during the generative period; TF5, sum of temperatures > 10°C from germination to full maturity; PF5, sum of precipitation from germination to full maturity.

Table 3. Correlation coefficients for quantitative variables.

Groups of traits	Traits	Correlation coefficient	
		Dimension 1	Dimension 2
Yield	Seed yield per plant, g (P1)	-0.46	0.39
	Weight of 100 seeds, g (P2)	-0.47	0.64
	Seed yield per area, g/m2 (P3)	-0.82	0.16
Morphology	Plant height, cm (M1)	0.15	0.93
	Height of the lowest pod, cm (M2)	0.10	0.90
Phenology	Days from germination to 50% flowering (F2)	0.84	0.16
	Days from germination to full maturity (F5)	0.97	0.13
	Days from 50% flowering to full maturity (F8)	0.91	0.11
Meteo	Sum of temperatures > 10°C, during the vegetative period, °C (TF2)	0.21	0.05
	Sum of precipitation during the vegetative period, mm (PF2)	0.80	-0.01
	Sum of temperatures > 10°C, during the period from germination to full maturity, °C (TF5)	0.82	0.18
	Sum of precipitation during the period from germination to full maturity, mm PF5	0.87	-0.08
	Sum of temperatures > 10°C during the generative period, °C (TF8)	0.79	0.18
	Sum of precipitation during the generative period, mm (PF8)	0.66	-0.13

GGE-biplot analysis

Yield remained the most informative trait reflecting genotype adaptability. Given the importance of yield as the primary economic trait, this variable was selected for genotype evaluation using the GGE-biplot method. The method followed the framework of Rakshit *et al* (2012), comparing genotypes relative to the 'ideal genotype', defined as maximum yield across environments. Multi-location data were analyzed without scaling ('Scaling = 0'), centred by environment ('Centering = 2'), and square-root transformed before analysis. Genotype evaluation employed genotype-centred singular value partitioning (SVP = 1) in the 'Mean vs. Stability' view.

According to the GGE-biplot framework, G × E interaction was represented by the first two principal components (PC1 and PC2). PC1 reflected the average yield of a genotype across all environments, whereas PC2 described its stability (variability across environments). In the present study, PC1 and PC2 together explained from 82.73% (PC1, 58.2%; PC2, 24.5%, Elitne) to 94% (PC1, 84.3%; PC2, 9.7%, Bokhonyky) of the total variation, validating the effectiveness of the model for genotype comparison (Figure 5).

In Elitne, the genotypes closest to the ideal genotype were 2083, 2085, 2087, and 2090. The standard-accessions Krasnokutskyi 123 and Rozanna were positioned at the periphery. Among the years, 2019 was closest to optimal conditions, whereas 2020 and 2021 deviated substantially. In Ustymivka, the most favourable year was 2020, while 2019 was the least favourable. The best-performing genotypes were 2085, 2087, 2076, and 2074. The standard Odisei showed low productivity across all years. Environmental conditions in Bokhonyky were unfavourable in all years, as indicated by the consistently negative PC1 direction. Years 2020 and 2021 were similar, whereas 2019 differed but remained suboptimal. The best genotypes were 2072, 2067, 2068, 2069, and 2065. Standards Rozanna and Krasnokutskyi 123 also performed well; Pamiat showed average performance.

The overall graph for three locations over three years (Figure 6) illustrates how large the difference between test points is and how difficult it is to find a universal ideal genotype: the vast majority of genotypes are located outside the circles. The genotype 2068 is highlighted as the closest to the ideal.

Adding the fourth extremely arid location Khlibodarske to the calculations slightly changed the overall axis loading indicators (PC1, 63.02%; PC2, 20.53%) (Figure 7). The Khlibodarske point, due to its low average yield and variability, has a low impact on the overall assessment and is located almost at the centre of the coordinates, but complements the overall assessment.

The SVP = 3 estimation method (singular value is symmetrically partitioned into the genotype and the environment eigenvectors) was applied to the same conditions, which allows us to assess 'which-won-where' (Figure 7B). The genotypes are located at the vertices of the polygon depending on their sensitivity to the environment. Thus, genotype 2068 reacts positively to the environment, and Odisei negatively. Genotype 2078 – closest to the centre of coordinates – is the most stable in all locations. This method also allows us to visualize accessions adapted to each location. For example, 2072 for Bokhonyky, 2085 and 2087 for Ustymivka and Elitne. As confirmed by the results of the analysis of each location separately (Figure 5). Genotype 2090 and Odisei have high yield potential, but are sensitive to growing conditions (Yan *et al*, 2007; Olivoto, 2023).

Due to the significant variability of weather conditions within each location, it was decided to analyze each location in each year as a single test site (Figure 8).

The conditions in Ustymivka in 2020 were optimal for the studied chickpea accessions. Genotypes 2068, 2083, 2087, and 2085 showed wide adaptability.

For additional validation of the stability of chickpea accessions, a Finley-Wilkinson regression analysis was performed (Figure 9).

Genotypes above the dashed line are sensitive to environmental conditions and have high yield potential, which is realized under favourable conditions. All standard-accessions are stable and are located below the dashed line.

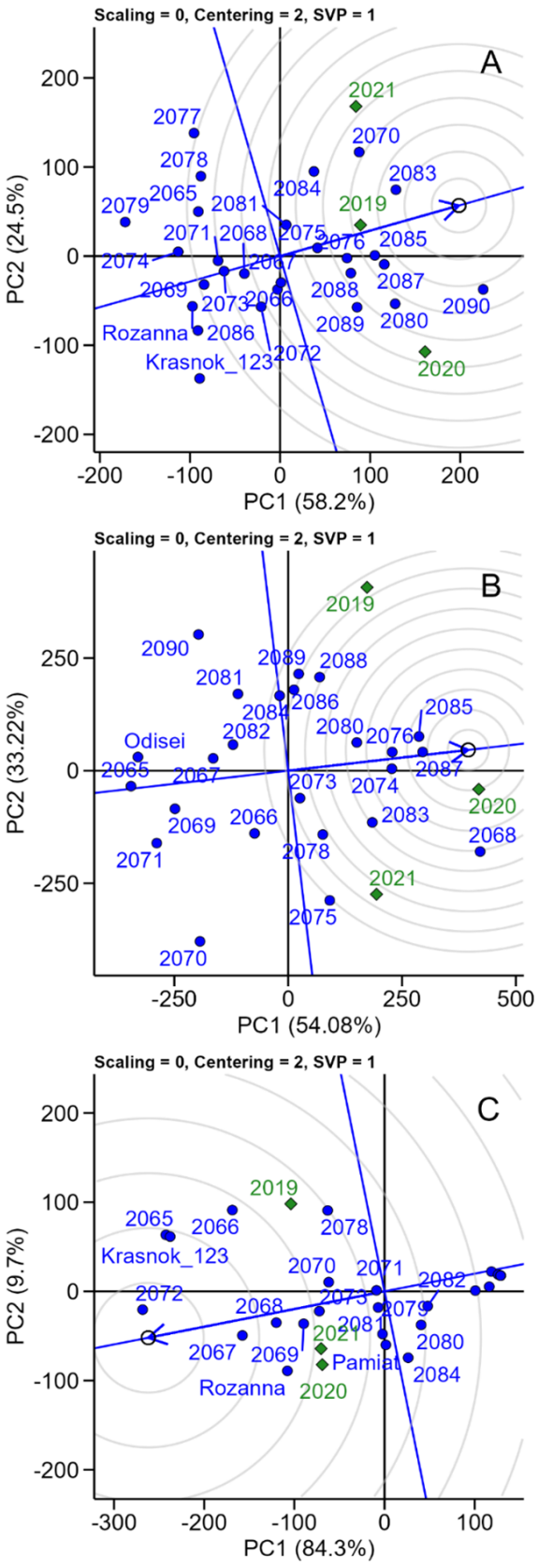


Figure 5. Ranking of genotypes (blue points) relative to the 'ideal genotype' (small circle at the average environment coordinate, AEC) for chickpea yield in three locations over three years (2019–2021; green diamonds): A, Elitne; B, Ustymivka; C, Bokhonyky.

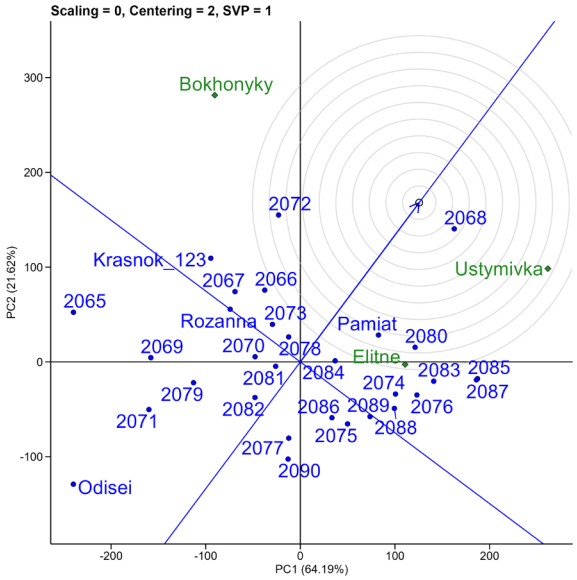


Figure 6. Ranking of genotypes (blue points) relative to the ideal genotype (small circle at the average environment coordinate, AEC) for chickpea yield in three locations (green diamonds), total for three years (2019–2021).

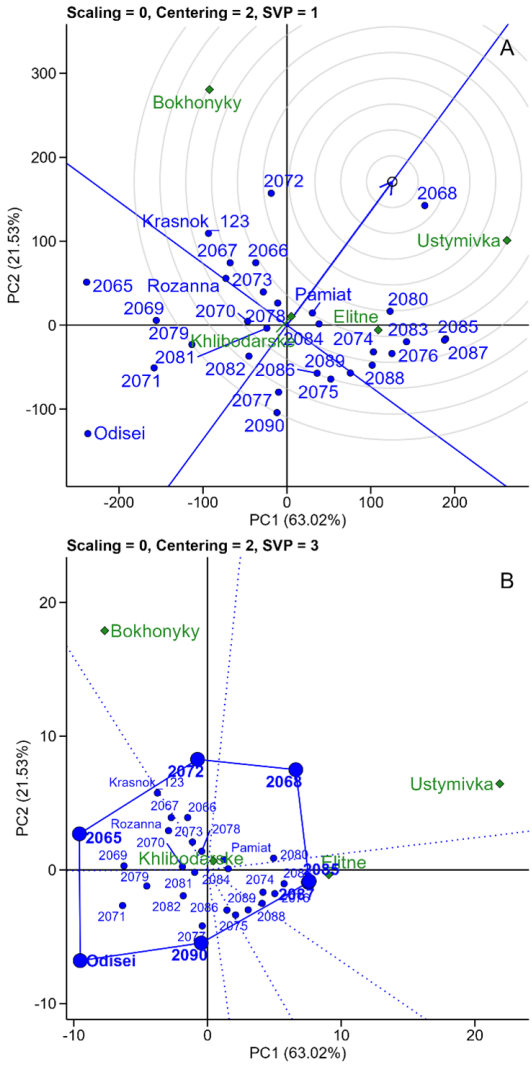


Figure 7. Ranking of genotypes (blue points) for chickpea yield in four locations (green diamonds), total for three years (2019–2021). A, relative to the ideal genotype (small circle at the average environment coordinate, AEC), B, 'which-won-where', i.e. the best adaptive genotypes for each location.

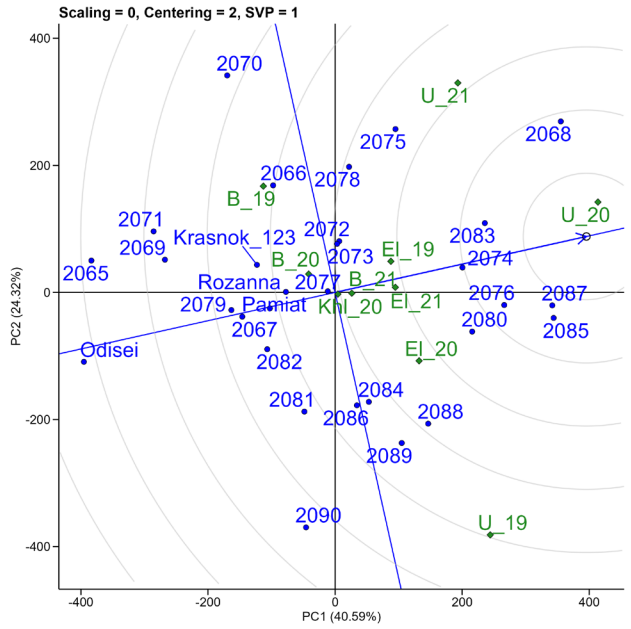


Figure 8. Ranking of genotypes (blue points) relative to an ideal genotype (the small circle on average environment coordinate for chickpea yield at each location by one year as an individual environment. Green diamonds indicate location × year combinations: El_19, El_20, El_21, Elitne; U_19, U_20, U_21, Ustymivka; B_19, B_20, B_21, Bokhonyky in 2019, 2020, 2021, respectively; Khl_20, Khlibodarske in 2020.

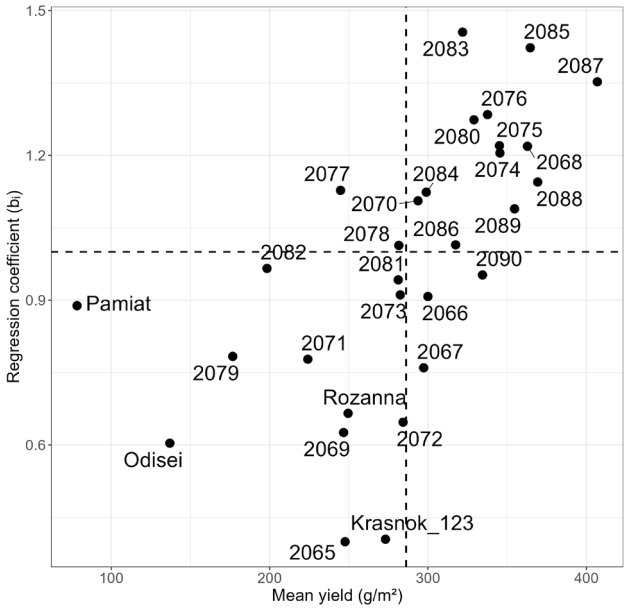


Figure 9. Finlay-Wilkinson (1963) regression plot: mean yield and environmental responsiveness (b_i) of genotypes across all environments. Dashed lines, average stability (b_i = 1) and mean yield of all accessions in this study (286.2g/m2)

Broadly adapted genotypes were defined as those showing regression coefficients close to unity ($b_i \approx 1$) combined with low deviations from regression (SD residuals), indicating average environmental responsiveness and high yield stability. Environments were defined as year × location combinations. Finlay-Wilkinson regression was performed using environmental indices calculated as the mean yield of all genotypes within each environment. Regression coefficients were used to estimate environmental responsiveness, while the standard deviation of residuals represented yield stability across environments.

For each of the analysis options, the genotypes were ranked: for GGE-biplot, Euclidean distance from this ideal genotype (Yan *et al*, 2007), for Finlay-Wilkinson (1963), accessions with a regression coefficient close to unity, low deviation from regression and high average yield (more than mean yield of all accessions on this research, 286.2g/m^2). The final ranking of genotypes was carried out according to the average rank of all options and allowed a better characterization of the studied genotypes according to their response to environmental conditions.

A detailed characterization of the most promising genotypes – adapted regionally or broadly – is presented in Table 4.

Discussion

The cultivation of chickpea in Ukraine is traditionally concentrated in the southern steppe regions and remains limited in the central and northern parts of the country. Introducing chickpea into crop rotations in non-traditional areas, such as the Western Forest-Steppe, represents a promising strategy for mitigating the impacts of climate change (Chernik & Tryhuba, 2023). The primary breeding centre for chickpea in Ukraine, PBGI in Odesa, has developed several drought-tolerant varieties, though these remain susceptible to diseases uncommon in their region of origin (Bushulian *et al*, 2018). The Forest-Steppe – a promising expansion zone for chickpea production – experiences both drought episodes and periods of excessive rainfall, combined with wide temperature fluctuations. These factors promote the development of fungal diseases, particularly Ascochyta blight, resulting in considerable yield losses. Therefore, new chickpea varieties must combine resistance to complex stress: drought, hyper-humidity and disease. Initial trials in the Right-Bank Forest-Steppe demonstrated that chickpea cultivation can be successful in favourable years, with yields comparable to those in the traditional southern cultivation zone (Bushulian *et al*, 2018; Makarchuk, 2021).

A 3-year programme of ecological testing across contrasting climatic zones enabled a comprehensive assessment of yield adaptability in selected chickpea breeding lines. Despite being located along similar latitudes, the three main testing sites – Elitne, Ustymivka and Bokhonyky – differed sharply in weather patterns. Waterlogging emerged as a critical stress factor, capable of inducing disease and prolonging the growing season through repeated vegetative growth. Because chickpea is an indeterminate species, excessive moisture may cause the abortion of pods and seeds, leading to partial or complete crop failure (Worku, 2016). According to Worku (2016), Kabuli-type chickpeas – represented among the genotypes in our study – are particularly sensitive to waterlogging.

Table 4. Characteristics of the most perspective genotypes of chickpea.

Genotype	Traits	Yield, g/m²						
		Optimal		Extremum		Yield (Mean)	Plasticity (Regression coefficient)	Stability (SD residuals)
		Elitne	Ustymivka	Bokhonyky (humidity)	Khlibodarske (drought)			
2068	Adaptability, drought resistance	387.7	650.0	32.6	73.8	362.8	1.22	132.15
2080	Adaptability, drought resistance	477.8	533.3	56.3	87.8	329.0	1.27	48.25
2085	Adaptability, drought resistance	476.9	590.0	2.8	77.5	364.6	1.42	75.84
2072	Hyper-humidity, stability	401.9	-	228.9	97.5	284.3	0.65	7.63
2087	Adaptability, drought resistance	472.4	590.0	0.0	68.3	406.9	1.35	86.60
2088	Adaptability, drought resistance	455.7	498.3	0.0	92.1	369.3	1.14	87.50
2076	Adaptability, plasticity	457.1	526.7	5.1	76.9	337.6	1.28	92.57
2083	Adaptability, drought resistance	494.7	535.3	17.8	74.7	321.8	1.46	52.31
2084	Stress tolerance, plasticity	447.9	454.0	68.6	77.6	298.9	1.12	73.17
2078	Adaptability, hyper-humidity	382.8	436.7	100.7	54.9	281.6	1.01	93.27
2067	Hyper-humidity	406.9	385.7	169.4	86.7	297.3	0.76	84.42
2065	Hyper-humidity	375.5	216.7	205.1	84.8	247.7	0.40	98.48
2069	Drought resistance	364.4	288.3	129.3	121.0	246.7	0.63	60.80
2075	High yield	440.7	448.3	0.0	93.7	345.1	1.22	117.51

Given the substantial yield losses associated with waterlogging, Dron *et al* (2022) emphasize developing new adapted varieties as the most effective mitigation strategy. Our multivariate analysis confirmed a consistent, negative response to adverse weather conditions across all yield-related traits. Since seed yield per unit area is of primary importance in production, this parameter was used to select the most adapted genotypes through GGE-biplot analysis.

The results showed that although the mean yield parameters in Elitne and Ustymivka were similar, yield variability in Ustymivka was considerably greater. This suggests a uniform yield response in the region of origin, but distinct mechanisms of environmental adaptation.

Across the different designs of assessing the adaptability of chickpea genotypes, GGE-biplot analysis showed that the first two principal components explained 64.92%–94% of the total variation, with PC1, which reflected the average yield of the genotype in all environments, accounting for 40.59%–84.3% and PC2, which described its variability in these environments, accounting for 9.7%–33.22%. These results align with previous findings (Farshadfar *et al*, 2013). This makes it possible to deeply and comprehensively characterize the studied genetic material on the adaptability of chickpea accessions to different growing conditions. Although chickpea is recognized as a drought-tolerant crop, terminal drought remins a major abiotic factor reducing yield. Climate change has shifted periods of extreme heat and drought beyond the southern regions into the Forest-Steppe (Elitne and Ustymivka). To broaden the assessment of adaptability, an extremely arid site – Khlibodarske (2020) – was added. A combined model using ten environments (three years across three sites, plus the ultra-arid site) highlighted the importance of both PC1 (40,59%) and PC2 (24,32%), supporting the use of such models for pattern extraction and noise reduction (Farshadfar *et al*, 2013).

To differentiate genotypes based on adaptability, we employed an ideal genotype comparison model like other authors (Yan & Rajcan, 2002; Yan & Tinker, 2006; Segherloo *et al*, 2010). The ideal genotype is conceptualized as having both high mean yield and high stability. Although no real genotype fully meets this criterion, it serves as a reference point. Each variant of the analysis provided information on the adaptability of the genotypes to environmental conditions, and the average rank allowed the characterization of the accessions according to the range of adaptability and stability. The first ten genotypes were characterized as highly adaptive (Table 4). Also, accessions with specific characteristics were distinguished, such as resistance to waterlogging (2065) or drought (2069). Analysis using the regression coefficient according to Finley-Wilkinson (1963) demonstrated the ability of the genotypes to respond to environmental challenges and added information on stability and plasticity. Three-

year testing across three locations revealed notable shifts in weather conditions and demonstrated the urgent need for new chickpea varieties. This was evident from the poor performance of established standard accessions, which were located far from the ideal genotype in Elitne and Ustymivka. However, under waterlogged conditions in Bokhonyky, both standards (Rozanna and Krasnokutskyi 123) performed near the ideal genotype.

Genotypes 2080 and 2085, among the five top-performing genotypes, have Krasnokutskyi 123 in their crossing combination, a variety that shows stable but moderate yields in the south; however, in the Forest-Steppe it performs with both higher stability and productivity (Bushulian *et al*, 2018). Genotypes 2087 and 2090, which were evaluated as highly adaptive to Ustymivka and Elitne conditions, originate from the hybrid combination Luhanets/Antey; both parents are productive but unstable and susceptible to Ascochyta. Antey is known for low yield stability and susceptibility to stress in its region of origin, Odesa (Bushulian *et al*, 2018).

Studies conducted by local scientists have emphasized the need to develop new, adapted varieties, which prompted the choice of this region for ecological testing of our chickpea lines (Gan *et al*, 2009; Zaparniuk & Sherepitko, 2011; Zaparniuk & Sherepitko, 2012). Genotypes selected as adaptive for Bokhonyky are really important for the region and represent potential sources of waterlogging tolerance. Accessions 2085, 2083, and 2087 were developed using donor material previously recommended for combining valuable traits (Vus *et al*, 2020).

Genotype 2068, which had the highest average adaptability rank, was registered with catalogue number UD0502245 at NCPGRU as a source of complex valuable traits under Certificates No. 2519 issued on 28 March 2024 (Sylenko *et al*, 2024a). Genotype 2087 was officially registered at NCPGRU as an adaptability source (Certificate N2612, accession name CIR 289-18, UD0502261; Vus, 2024). Genotypes 2075 (UD0502251) and 2088 (UD0502262) were registered at NCPGRU as sources of complex valuable traits under Certificates No. 2520 and 2521, issued on 28 March 2024 (Sylenko *et al*, 2024b and 2024c).

Conclusions

The present study provides an ecological evaluation of chickpea genetic resources under novel and contrasting environmental conditions relevant to ongoing climate change. Adaptability and stability of accessions were assessed in the Ukrainian Forest-Steppe, as well as under extreme moisture regimes, including extra-arid and hyper-humid environments. The evaluation revealed substantial variation in adaptive responses among genotypes and enabled the identification of accessions with stable performance across environments, including newly determined genetic resources with potential tolerance to waterlogging.

Overall, the results expand the characterization of chickpea germplasm by documenting adaptive traits expressed under non-standard and climatically challenging conditions. The generated data contribute to a better understanding of the adaptive capacity of chickpea genetic resources and provide a valuable reference for their informed use in future research and breeding activities aimed at developing climate-resilient cropping systems.

Supplemental data

Supplemental Table 1. Climatic and weather characteristics of the study locations

Author contributions

Nadiia Vus: conceptualization, data curation, formal analysis, investigation, methodology, resources, software, supervision, validation, writing – original draft, writing – review & editing. Olha Bezuhla: Conceptualization, data curation, formal analysis, funding acquisition, investigation, methodology, resources, supervision, validation, writing – original draft, writing – review & editing. Serhii Sylenko: data curation, investigation, methodology, project administration, resources, software, validation, writing – original draft, Writing – review & editing. Antonina Vasylenko: conceptualization, formal analysis, investigation, methodology, validation, writing – original draft, writing – review & editing. Viacheslav Sichkar: data curation. writing – original draft, writing – review & editing. Mykola Kondratenko: data curation, formal analysis, investigation, methodology, resources, validation, writing – original draft, writing – review & editing. Margarita Barylko: data curation. writing – original draft, writing – review & editing

Conflict of interest statement

The authors have declared that no competing interests exist.

References

Berrada, A.F., Shivakumar, B.G., Yaduraju, N.T. (2007). Chickpea in cropping systems. In: Yadav S.S., Redden R.J., Chen W. & Sharma B., eds. Chickpea Breeding and Management. Pondicherry, India: CAB. International, 193-212.

Bushulian, O.V., Stelmach, A.F., Lamari, N.P., Fait, V.I. (2018). Genotype reaction and stability of crop production of chickpea (*Cicer arietinum* L.) in years of testing. F*actors of Experimental Evolution of Organisms*, 23, 35-39. (In Ukrainian).

Chernik, I. V., Tryhuba, O. V. (2023). Common chickpea (Cicer arietinum L.) – a prospective legume culture of the Western Forest Steppe of Ukraine. *Scient. not. Ternop. national. pedagogical. un. Ser. Biol* (Nauk. Zap. Ternop. nat. ped. univ. Ser. Biol.), 83(3–4), 118-126. (In Ukrainian).

Dron, N., Simpfendorfer, S., Sutton, T., Pengilley, G., Hobson, K. (2022). Cause of Death: Phytophthora or Flood? Effects of Waterlogging on Phytophthora medicaginis and Resistance of Chickpea (*Cicer arietinum*). *Agronomy*, 12(1), 89. https://doi.org/10.3390/agronomy12010089

FAOSTAT (2023). https://www.fao.org/faostat/en/#data/QCL [Accessed 07 mai 2025].

Farshadfar, E., Rashidi, M., Jowkar, M., Zali, H. (2013). GGE Biplot analysis of genotype × environment interaction in chickpea genotypes. *Europ. J.l of Exp. Biol.*, 3, 417–423.

Finlay, K.W., Wilkinson, G.N. (1963). The Analysis of Adaptation in a Plant-Breeding Programme. *Australian Journal of Agricultural Research*, 14 (6). CSIRO PUBLISHING: 742. https://doi.org/10.1071/AR9630742

Fritsche-Neto, R., Resende, R. T., Olivoto, T., Garcia-Abadillo, J., Nascimento, M., Bahia, M. A. M., Jarquin, D., & Vieira, R. A. (2025). Prediction-based breeding: Modern tools to

optimize and reshape programs. *Crop Science*, 65, e70175. https://doi.org/10.1002/csc2.70175

Gan, Y., Zentner, R. P., McDonald, C. L., Warkentin, T., Vandenberg, A. (2009). Adaptability of chickpea in northern high latitude areas—Maturity responses. *Agric. For. Meteor.*, 149, 711-720. https://doi.org/10.1016/j.agrformet.2008.10.026

Gaur, P.M., Gowda, C.L.L., Knights, E.J., Warkentin, T., Açikgoz, N., Yadav, S. S., Kumar, J. (2007). Breeding achievements. In: Yadav, S.S., Redden, R.J., Chen, W., Sharma, B., eds. Chickpea Breeding and Management. Pondicherry, India: CAB. International, 391-416.

Genesys-PGR. International Plant Genetic Resources for Food and Agriculture (PGRFA) information platform. https://www.genesys-pgr.org [Accessed 26.11.2025].

Henry, L., Wickham, H. (2023). rlang: Functions for Base Types and Core R and 'Tidyverse' Features. R package version 1.1.1. https://CRAN.R-project.org/package=rlang

Hijmans, R. (2025). terra: Spatial Data Analysis. R package version 1.8-60. https://CRAN.R-project.org/package=terra

Hristov, J., Toreti, A., Pérez Domínguez, I., Dentener, F., Fellmann, T., Elleby, C., Ceglar, A., Fumagalli, D., Niemeyer, S., Cerrani, I., Panarello, L., Bratu, M. (2020). Analysis of climate change impacts on EU agriculture by 2050: JRC PESETA IV project: Task 3. Publications Office https://publications.jrc.ec.europa.eu/repository/handle/JRC119632

Khan, M. N., Wahab, S., Wahid, N., Shah, S. N., Ullah, B., Kaplan, A., Razzaq, A., Bibi, M., Suleman, F., Ali, B. (2024). Chapter 4 - Impact of climate change on yield and quality of legumes. In: Fahad S., Adnan M., Munir I., Lal R., Nawaz T. & Shah S., eds. Challenges and Solutions of Climate Impact on Agriculture. Academic Press, 85-111. https://doi.org/10.1016/B978-0-443-23707-2.00004-0

Kobyzeva, L. N., Bezugla, O. M., Silenko, S. I., Kolotilov, V. V., Sokol, T. V., Dokukina, K. I., Vasylenko, A. O., Bezugliy, I. M., Vus, N. (2016). Methodical recommendations for the study of genetic resources of leguminous crops. Kharkiv: NAAS, Plant Production Institute V. Ya. Yuryev. (In Ukrainian).

Lampkin, N.H., Pearce, B.D., Leake, A.R., Creissen, H., Gerrard, C.L., Girling, R., Lloyd, S., Padel, S., Smith, J., Smith, L.G., Vieweger, A., Wolfe, M.S. (2015). The role of agroecology in sustainable intensification. Report for the Land Use Policy Group. Organic Research Centre, Elm Farm and Game & Wildlife Conservation Trust. 163 p

Makarchuk, M. A. (2021). The economically and valuable feature of selection number of chickpea (*Cicer arietinum* L.) in the condition of the Right-Bank Forest-Steppe. *Journal of Uman NUH*, 98(1), 210-219. https://doi.org/10.31395/2415-8240-2021-98-1-210-219

Mohammadi, S., Rydgren, K., Bakkestuen, V., Gellespie, M.A. (2023). Impacts of recent climate change on crop yield can depend on local conditions in climatically diverse regions of Norway. *Scientific Reports*, 13, 3633. https://doi.org/10.1038/s41598-023-30813-7

Olivoto, T. (2023). Analyzing multienvironment trials using GGE. https://tiagoolivoto.github.io/metan/articles/vignettes_gge.html

Olivoto, T., Lúcio, A. D. (2020). metan: an R package for multi-environment trial analysis. *Meth. in Ecol. and Evol.*, 11, 783–789. https://doi.org/10.1111/2041-210X.13384

Pedersen, T. (2024). patchwork: The Composer of Plots. R package version 1.3.0. https://CRAN.R-project.org/package=patchwork

Pixley, K. V., Cairns, J. E., Lopez-Ridaura, S., Ojiewo, C. O., Dawud, M. A., Drabo, I., Mindaye, T., Nebie, B., Asea, G., Das, B., Daudi, H., Desmae, H., Batieno, B. J., Boukar, O., Mukankusi, C.T.M., Nkalubo, S.T., Hearne, S. J., Dhugga, K. S., Gandhi, H., Snapp, S., Zepeda-Villarreal, E. A. (2023). Redesigning crop varieties to win the race between climate change and food security. *Molecular Plant*, 16(10), 1590-1611. https://doi.org/10.1016/j.molp.2023.09.003.

Rakshit, S., Ganapathy, K.N., Gomashe, S.S., Rathore, A., Ghorade, R., Kumar, N., Ganesmurthy, K., Jain, Sk., Kamatar, M., Sachan, J., Ambekar, S., Ranwa, B., Kanawade, D., Balusamy, M., Kadam, D., Sarkar, A., Tonapi, V., Patil, J. (2012). GGE biplot analysis to evaluate genotype, environment and their interactions in sorghum multi-location data. *Euphytica*, 185, 465–479 https://doi.org/10.1007/s10681-012-0648-6

Rani Das, R., Laporte, M.-A., Valette, L., Hazekamp, T., Arnaud, E., Rathore, A. (2024). Planteome/CO_338-chickpea-traits: CO_338-chickpea-traits ontology (V1.0) [Data set]. Zenodo. https://doi.org/10.5281/zenodo.11471775 / https://cropontology.org/term/CO_338:ROOT.

Raza, A., Razzaq, A., Mehmood, S.S., Zou, X., Zhang, X., Lv, Y., Xu, J. (2019). Impact of climate change on crops adaptation and strategies to tackle its outcome: A Review. *Plants* (Basel), 8(2), 34. https://doi.org/10.3390/plants8020034

R Core Team (2023). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Rebollo, I., Aguilar, I., Pérez de Vida, F., Molina, F., Gutiérrez, L., & Rosas, J. E. (2023). Genotype by environment interaction characterization and its modeling with random regression to climatic variables in two rice breeding populations. *Crop Science*, 63, 2220–2240. https://doi.org/10.1002/csc2.21029

Segherloo, A. E., Sabaghpour, S. H., Dehghani, H., Kamrani, M. (2010). Screening of superior chickpea genotypes for various environments of Iran using genotype plus genotype × environment (GGE) biplot analysis. *J. of Plant Breed. and Crop Sci.*, 2(9), 286–292. https://doi.org/10.5897/JPBCS.9000115

Schauberger, P., Walker, A. (2022). openxlsx: Read, Write and Edit xlsx Files_. R package version 4.2.5.1. https://CRAN.R-project.org/package=openxlsx

Sloat, L.L., Davis, S.J., Gerber, J.S., Moore, F. C., Ray, D. K., West, P. C., Mueller, N. D. (2020). Climate adaptation by crop migration. *Nature Communication*, 11, 1243. https://doi.org/10.1038/s41467-020-15076-4

Slowikowski, K. (2024). ggrepel: Automatically Position Non-Overlapping Text Labels with 'ggplot2'. R package version 0.9.6. https://CRAN.R-project.org/package=ggrepel.

Sylenko, S., Vus, N., Andrushchenko, O., Bezugla, O., Kobyzeva, L. (2024a). Certificate No. 2519 of registration of accession of plant gene pool. *Cicer arietinum*. L 261-18, UD0502245. Chickpea. Application No. 0050096. 23.03.2023. Authors: Sylenko S., Vus N., Andrushchenko O., Bezugla O., Kobyzeva L. Kharkiv, Ukraine. National Center of Plant Genetic Resources of Ukraine (in Ukrainian).

Sylenko, S., Vus, N., Andrushchenko, O., Bezugla O., Kobyzeva L. (2024b). Certificate No. 2520 of registration of accession of plant gene pool. Cicer arietinum. L 272-18, UD0502251. Chickpea. Application No. 0050097. 23.03.2023. Authors: Sylenko S., Vus N., Andrushchenko O., Bezugla O., Kobyzeva L. 28.03.2024. Kharkiv, Ukraine. National Center of Plant Genetic Resources of Ukraine (in Ukrainian).

Sylenko, S., Vus, N., Andrushchenko, O., Bezugla, O., Kobyzeva, L. (2024c). Certificate No. 2521 of registration of accession of plant gene pool. *Cicer arietinum*. L 290-

18, UD0502262. Chickpea. Application No. 0050098. 23.03.2023. Authors: Sylenko S., Vus N., Andrushchenko O., Bezugla O., Kobyzeva L. 28.03.2024. Kharkiv, Ukraine. National Center of Plant Genetic Resources of Ukraine (in Ukrainian).

The Plant Breeding and Genetics Institute – National Center of Seed and Cultivar Investigation, 2024. Official page. https://sgi.in.ua/ [Accessed 07 May 2025].

Vaissie, P., Monge, A., Husson, F. (2023). Factoshiny: Perform Factorial Analysis from 'FactoMineR' with a Shiny Application. R package version 2.5. https://CRAN.R-project.org/package=Factoshiny

Varshney, R. K., Roorkiwal, M., Sun, S., Bajaj, P., Chitikineni, A., Thudi, M., Singh, N. P., Du, X., Upadhyaya, H. D., Khan, A. W., Wang, Y., Garg, V., Fan, G., Cowling, W. A., Crossa, J., Gentzbittel, L., Voss-Fels, K. P., Valluri, V. K., Sinha, P., Singh, V. K., Liu, X. (2021). A chickpea genetic variation map based on the sequencing of 3,366 genomes. *Nature*, 599(7886), 622–627. https://doi.org/10.1038/s41586-021-04066-1

Vus, N., Kobyzeva, L. N., Bezuglaya, O. N. (2020). Determination of the breeding value of collection chickpea (Cicer arietinum L.) accessions by cluster analysis. *Vavilov J. of Genetics and Breed.*, 24(3), 244–251. https://doi.org/10.18699/VJ20.617

Vus, N. (2024). Certificate No. 2612 of registration of accession of plant gene pool. *Cicer arietinum* L. CIR 289-18, UD0502261. Chickpea. Application No. 0053000. 29.10.2024. Author: Vus N. Collectors: Sylenko S., Sichkar V, Bezugla O. 25.11.2024. Kharkiv, Ukraine. National Center of Plant Genetic Resources of Ukraine.

Watson, C. A., Reckling, M., Preissel, S., Bachinger, J., Bergkvist, G., Kuhlman,T., Lindström, K., Nemecek, T., Topp, C. F.E., Vanhatalo, A., Zander, P., Murphy-Bokern, D., Stoddard, F. L. (2017). Chapter 4. - Grain Legume Production and Use in European Agricultural Systems. In: Sparks D. L., eds. Advances in Agronomy. Academic Press, 144, 235-303. https://doi.org/10.1016/bs.agron.2017.03.003

von Wettberg, E. J. B., Chang, P. L., Başdemir, F., Carrasquila-Garcia, N., Korbu, L. B., Moenga, S. M., Bedada, G., Greenlon, A., Moriuchi, K. S., Singh, V., Cordeiro, M. A., Noujdina, N. V., Dinegde, K. N., Shah Sani, S. G. A., Getahun, T., Vance, L., Bergmann, E., Lindsay, D., Mamo, B. E., Warschefsky, E. J., Dacosta-Calheiros, E., Marques, E., Yilmaz, M. A., Cakmak, A., Rose, J., Migneault, A., Krieg, C. P., Saylak, S., Temel, H., Friesen, M. L., Siler, E., Akhmetov, Z., Ozcelik, H., Kholova, J., Can, C., Gaur, P., Yildirim, M., Sharma, H., Vadez, V., Tesfaye, K., Woldemedhin, A.F., Tar'an, B., Aydogan, A., Bukun, B., Penmetsa, R. V., Berger, J., Kahraman, A., Nuzhdin, S. V., Cook, D. R. (2018). Ecology and genomics of an important crop wild relative as a prelude to agricultural innovation. *Nat. communic.*, 9(1), 649. https://doi.org/10.1038/s41467-018-02867-z

Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis. New York: Springer-Verlag.

Worku, W. (2016). Waterlogging effects on growth, nodulation and productivity of Desi and Kabuli chickpea (Cicer arietinum L.). *Ethiopian Journal of Biological Sciences*, 15(1), 55–77.

Yan, W., Kang, M. S., Ma, B., Woods, S., Cornelius, P. L. (2007). "GGE Biplot vs. AMMI Analysis of Genotype-by-Environment Data". *Crop Science*, 47 (2): 641–53. https://doi.org/10.2135/cropsci2006.06.0374

Yan, W., Tinker, N. A. (2006). Biplot analysis of multi-environment trial data: Principles and applications. *Can. Journ. Plant Sci.*, 86(3), 623–645. https://doi.org/10.4141/P05-169

Yan, W., Rajcan, I. (2002). Biplot analysis of test sites and trait relations of soybean in Ontario. *Crop Sci.*, 42(1), 11–20. https://doi.org/10.2135/cropsci2002.1100

Zaparniuk, O.F., Sherepitko, V.V. (2011). Genotypic differences in the manifestation of morphobiological signs of chickpea plants (*Cicer arietinum*. L.). Collection of scientific works of Vinnytsia National Agrarian University (Zbirnyk naukovykh prats' Vinnyts'koho natsional'noho ahrarnoho universytetu), 7(47), 79-83. (In Ukrainian).

Zaparniuk, O.F., Sherepitko, V.V. (2012). Manifestation of economically valuable traits of chickpea plants (*Cicer arietinum* L.) under soil and climatic conditions of the Podolsk region. Collection of scientific papers of the Institute of Bioenergy Crops and Sugar Beet (Zbirnyk naukovykh prats' Instytutu bioenerhetychnykh kul'tur i tsukrovykh buryakiv), 14, 430-434. (In Ukrainian).

Genetic Resources

OTHER ARTICLES

Genetic Resources (2026), 7 (13), 143–152
DOI: 10.46265/genresj.KVVE1674
https://www.genresj.org
ISSN: 2708-3764

Rose genetic resources conserved *ex situ* at the M.M. Gryshko National Botanical Garden, Ukraine

Olena Rubtsova*, Mykola Shumyk, Nataliia Chuvikina, Tetyana Vakulenko, Valentina Chizhankova

M.M. Gryshko National Botanical Garden, National Academy of Sciences of Ukraine, 1 Sadovo-Botanichna Str., Kyiv 01103, Ukraine

Abstract: The rose genetic collection of the M.M. Gryshko National Botanical Garden, National Academy of Sciences of Ukraine (Kyiv, Ukraine) is the largest and most systematically studied *Rosa* collection in the country. Established in 1946 on the basis of planting material obtained from Germany as part of post-war reparations, it has expanded significantly over eight decades and today comprises 18 wild species and over 650 cultivars of various origins. The collection serves as a long-term *ex situ* conservation resource and provides a basis for studies of morphological traits, winter hardiness, drought tolerance and resistance to major pathogens. Monitoring has identified cultivars with high resistance to powdery mildew, black spot and rust, while morphological analyses support taxonomic and breeding research. The collection also includes historically important heritage cultivars and 12 original Ukrainian cultivars registered in the State Register of Plant Varieties Suitable for Distribution. Seeds of wild species are preserved in the seed laboratory, and a specialized seedbank is planned for long-term conservation. Extensive national and international collaborations contribute to the continuous enrichment of the gene pool and support educational, scientific and outreach activities. The study highlights the scientific, cultural and conservation value of this unique *Rosa* collection and its importance for breeding programmes in northern Ukraine.

Keywords: *Rosa*, genetic resources, *ex situ* conservation, rose breeding, Ukraine

Citation: Rubtsova, O., Shumyk, M., Chuvikina, N., Vakulenko, T. and Chizhankova, V. (2026) “Rose genetic resources conserved *ex situ* at the National Botanical Garden (Ukraine)”, *Genetic Resources*, 7(13), pp. 143–152. doi: 10.46265/genresj.KVVE1674.

Introduction

Roses (*Rosa* L.) represent one of the most diverse and economically important groups of ornamental plants worldwide. The diversity of wild species in the genus *Rosa* is the result of natural evolutionary processes, whereas its long history of cultivation and breeding has produced thousands of cultivars (Young and Schorr, 2007) that serve as sources of valuable traits, including winter hardiness, drought tolerance, disease resistance and ornamental qualities. Preserving this diversity is essential for sustainable breeding, adaptation to climate change, and the protection of horticultural heritage.

In this context, the historical development of garden roses provides an important framework for understanding the origin and diversity of cultivated forms. Garden roses have a long and complex history of cultivation, originating from wild species native to Europe, Asia and North America. Early domestication and selection began in ancient civilizations, particularly in China and the Middle East, where roses were valued for their ornamental, medicinal and symbolic properties. A major turning point in rose breeding occurred in the late 18th and early 19th centuries with the introduction of Chinese roses into Europe, bringing traits such as recurrent flowering and a broader colour range. This led to the development of new classes, including Hybrid Tea, Floribunda, and Polyantha roses, which dominate modern horticulture. Continuous breeding efforts throughout the 19th and 20th centuries have resulted in thousands of cultivars with diverse morphological and ornamental characteristics adapted to a

* Corresponding author: Olena Rubtsova (olenarubtsova@gmail.com)

Received: 25.01.2026 | Accepted: 02.04.2026 | Published online: 23.04.2026

wide range of climatic and landscape conditions.

Within this diversity, old garden roses are defined as cultivated classes that originated before the introduction of the first Hybrid Tea rose ('La France', 1867) (Krüssmann, 1981), which is widely accepted as the boundary between old and modern roses. These include Gallica, Damask, Alba, Centifolia, and Moss roses. In addition, the term heritage cultivars refers to historically significant cultivars that have been preserved in cultivation due to their cultural, genetic, or horticultural value, often representing earlier stages of breeding and no longer widely used in modern commercial production.

From a historical perspective, several key stages can be distinguished in European rose breeding: early cultivation and selection of European species and ancient garden groups (up to the 18th century); the introduction of Chinese roses into Europe in the late 18th century, enabling recurrent flowering; the development of hybrid groups such as Bourbon, Noisette, and Hybrid Perpetual roses in the 19th century; and finally, the emergence of modern roses following the introduction of Hybrid Tea roses from the late 19th century onward.

Ex situ conservation of rose genetic resources is predominantly carried out through living field collections, as clonal maintenance is required to preserve cultivar identity and phenotypic stability. Botanical gardens, therefore, play a crucial role in safeguarding rose germplasm, providing long-term maintenance, systematic documentation, and opportunities for characterization and evaluation under local environmental conditions. In contrast to seed-based conservation, living collections allow continuous observation of adaptive traits and enable the identification of valuable genotypes for breeding and landscaping (Collective, 2023; Lempitsky and Halytska, 1968).

The M.M. Gryshko National Botanical Garden, National Academy of Sciences of Ukraine (M.M. Gryshko NBG) maintains the largest and most comprehensive rose genetic collection in the country. The collection was established in 1946, following the introduction of planting material from Germany in the post-war period. The rose collection was introduced from nurseries located in Saxony (Germany), including Otto Kloß, Heinrich Tietze, Max Ziegenbalg, Paul Hauber, Karl Köhler, Victor Teschendorff, Emil Teich, Münch und Haufe, Guido Geißler, Theodor Simmgen, and the Rosarium Sangerhausen (Scientific Archive M.M. Gryshko NBG, 1946). and has since been continuously enriched through national and international exchanges (Gryshko, 1949; Ahr and Lang, 1991; Cherevchenko and Chuvikina, 2003) (Figure 1).

Over several decades, the collection has evolved from a display-oriented rose garden into a structured *ex situ* genetic resource, supporting research, breeding, and conservation objectives.

M.M. Gryshko NBG is located in the Right-Bank Forest-Steppe zone of Ukraine, characterized by a moderately continental climate with cold winters, occasional extreme frost events, and periodic summer droughts. These environmental conditions provide a natural framework for evaluating winter hardiness, drought tolerance and resistance to major fungal diseases in roses. Long-term observations of accessions under open-ground conditions (Figure 2a,c) have generated valuable data on phenotypic stability and adaptive potential, particularly relevant for northern and central regions of Eastern Europe (Rubtsova and Chizhankova, 2017).

In addition to its scientific and practical value, the rose collection of the M.M. Gryshko NBG has significant historical and cultural importance. It includes a substantial proportion of heritage cultivars representing key stages in European rose breeding, as well as original Ukrainian cultivars developed for local environmental conditions. Specifically, cultivars bred at the Nikita Botanical Garden ('Ahtiar', 'Emmi', 'Kakhovka', 'Khersones', 'Klimentina', 'Korallovyi Siurpriz', 'Krasnyi Maiak', 'Krymskii Samotsvet', 'Maikl', 'Pestraia Fantaziia', 'Plamja Vostoka', 'Polka Babochka', 'Professor Viktor Ivanov', 'Zolotaia Osen') as well as those developed at the M.M.

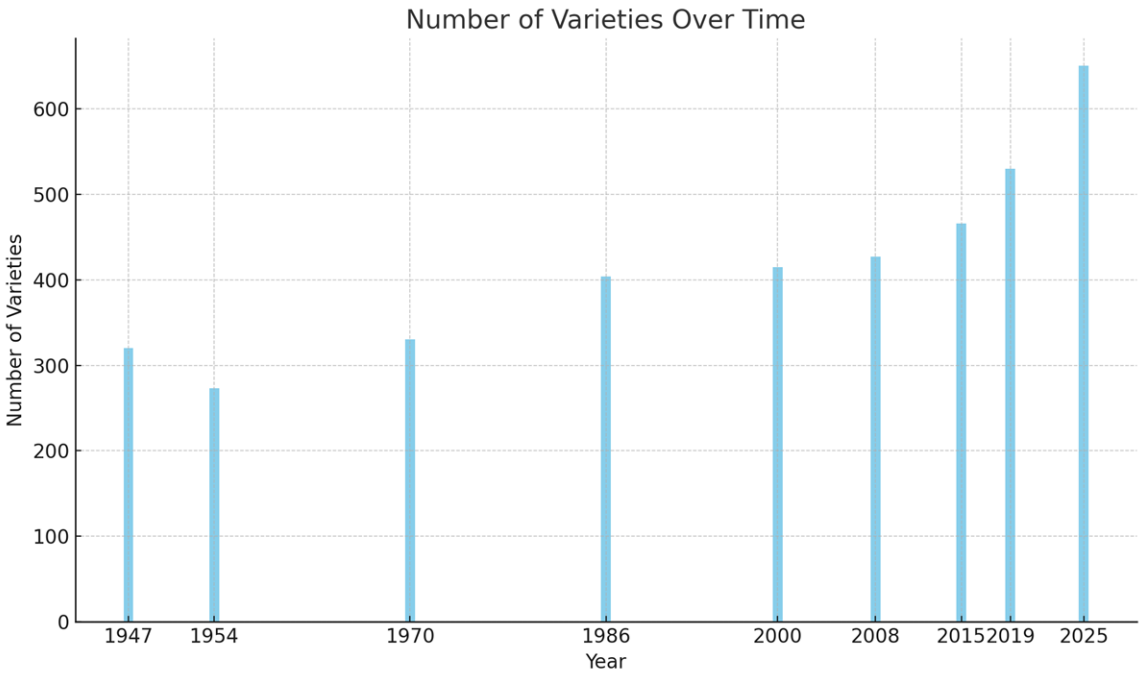


Figure 1. Number of rose varieties conserved over time at the M.M Gryshko National Botanical Garden, Ukraine.

Gryshko NBG (‘Akvarel Rose Park’, ‘Halaktyka’, ‘Khortytsia’ (Figure 2d) ‘Hratsiinyi Tanok’ (Figure 2e), ‘Karusel’, ‘Kraplia Sontsia’, ‘Mushli’, ‘Pochaina’, ‘Solodkyi Son’, ‘Spohady’, ‘Vintazh’, ‘Vrazhennia’). In recognition of its exceptional value, the collection was included in the State Register of Scientific Objects of National Heritage of Ukraine in 2006 (Rakhmetov *et al*, 2019).

This article presents an overview of the rose genetic resources conserved *ex situ* at the M.M. Gryshko NBG, focusing on the composition of the collection, conservation and management practices, characterization and evaluation of accessions, and the use of the collection in research, breeding, education and public outreach. By documenting this collection in a standardized format, the study contributes to the integration of Ukrainian ornamental plant genetic resources into the international framework of plant genetic resource conservation.

Collection description

The rose genetic collection of the M.M. Gryshko NBG represents a national reference collection of *Rosa* genetic resources, combining conservation, research and practical horticultural functions.

The germplasm is initially maintained in the nursery, where propagation is carried out by grafting (budding). Once a sufficient number of plants is obtained, they are transplanted to the rose garden: shrub roses are represented by at least three plants, which are planted together in the same location, while hybrid tea and floribunda cultivars are

Figure 2. The rose collection of the M.M. Gryshko NBG: history, conservation and utilization. (a), rose garden in 1948, representing the early stage of establishment of the rose collection; (b), air-dry shelter; (c), general view of the rose garden as an *ex situ* conservation site; (d), ‘Khortytsia’, a Ukrainian-bred rose cultivar; (e), ‘Hratsiinyi Tanok’, a Ukrainian-bred rose cultivar; (f), cabinet for storage of fruit and seed collections (world flora) at the M.M. Gryshko NBG; on the right, a pull-out drawer with specimens of the Rosaceae family (genus *Rosa*); (g), rose breeding course for student; (h), open-air concert in rose garden. Sources: Fig. 4a – Museum of the M.M. Gryshko NBG; Fig. 4b–e, g, h – photos by O. L. Rubtsova, Fig.4f – photo by T.B. Vakulenko.

planted in groups of up to 150 individuals.

Figure 1 illustrates the increase in the number of rose cultivars maintained in the collection from 1947 to 2025. An initial phase of modest growth is observed between 1947 and 1986, with the number of varieties rising from approximately 320 to 400. A more pronounced expansion occurred from 2000 onwards, reflecting intensified efforts in acquisition, international collaboration and propagation. By 2025, the collection reached over 650 cultivars, indicating a steady and substantial enhancement of genetic diversity within the rose collection.

Taxonomic and horticultural composition

At present, the collection comprises 18 wild *Rosa* species and more than 650 cultivated varieties, representing the main horticultural groups used in ornamental horticulture. The diversity of the collection reflects both the taxonomic breadth of the genus *Rosa* and the historical development of garden roses.

Cultivated accessions include representatives of modern garden rose groups such as Hybrid Tea, Floribunda, Grandiflora, Miniature, Shrub, Climbing, including both ornamental garden types and cultivars used for cut flowers, as well as a substantial group of old roses.

This wide representation allows comparative evaluation of genotypes with different growth habits, flowering patterns and adaptive traits under uniform environmental conditions.

Wild *Rosa* species

Wild *Rosa* species constitute an essential component of the collection, serving as sources of adaptive traits such as winter hardiness, drought tolerance, disease resistance and specific morphological characteristics. The collection (Table 1) includes species of Eurasian, East Asian and North American origin, several of which are of particular interest for breeding and conservation (Rubtsova *et al*, 2025b).

Table 1. Wild *Rosa* species conserved in the M.M. Gryshko rose collection: origin, main characteristics and number of preserved individuals

No.	*Rosa* species	Origin	Main characteristics	No. of individuals
1	*R. arkansana* Porter	North America	Stems erect, slender or stout, 0.6–1.5m tall; pink flowers	3
2	*R. canina* L.	Europe, Northwest Africa, and Western Asia	Shrub 1–5m tall; pale pink flowers	10
3	*R. caudata* Baker	China	Shrub up to 4m tall; red flowers, in corymbs, 3.5–6cm in diameter	2
4	*R. gallica* L.	Southern and Central Europe, eastwards to Turkey and the Caucasus	Shrub up to 1m tall; deep pink flowers	3
5	*R. glauca* Pourr.	Central and Southern Europe	Shrub 1.5–3m tall; glaucous blue-green to coppery or purplish leaves; pink flowers	3
6	*R. donetzica* Dubovik	Eastern Europe (Ukraine); rare species listed in the Red Book of Ukraine (2009)	Shrub up to 0.9m tall; pink flowers	3
7	*R. filipes* Rehder & E.H. Wilson	Western China	Shrub 3–5m tall; white flowers , 2–2.5cm in diameter, in large corymbs	3
8	*R. kokanica* (Regel) Regel ex Juz.	Afghanistan, Kazakhstan, Mongolia; Southwest Asia	Shrub 1.5–2m tall; pale yellow flowers	3
9	*R. multiflora* Thunb.	East Asia (China, Korea, Japan)	Climbing shrub up to 3–5m; small flowers (1.5–4cm), in large corymbs, white or pink	3
10	*R. nitida* Willd.	Northeastern North America	Shrub up to 1m tall; glossy leaves; pink flowers	3
11	*R. rubiginosa* L.	Europe and Western Asia	Shrub up to 2–3m tall; foliage with strong apple-like fragrance; pink flowers	3
12	*R. rugosa* Thunb.	East Asia	Shrub up to 1.5m tall; leaflets distinctly corrugated; dark pink to white flowers	3
13	*R. roxburghii* Tratt.	East Asia	Shrub up to 2.5m tall; pink flowers; large burred hips resembling chestnuts	3
14	*R. sambucina* Koidz.	East Asia	Climbing shrub; white flowers, in corymbs	3
15	*R. setigera* Michx.	Central and Eastern North America	Climbing shrub up to 3m; pink flowers	2
16	*R. spinosissima* L.	Western, Central, and Southern Europe and Northwest Africa	Shrub up to 1.5m tall; pale pink flowers	3
17	*R. virginiana* Mill.	Eastern North America	Shrub up to 2m tall; pink flowers	2
18	*R. xanthina* Lindl.	East Asia	Shrub up to 2.5m tall; yellow flowers	2

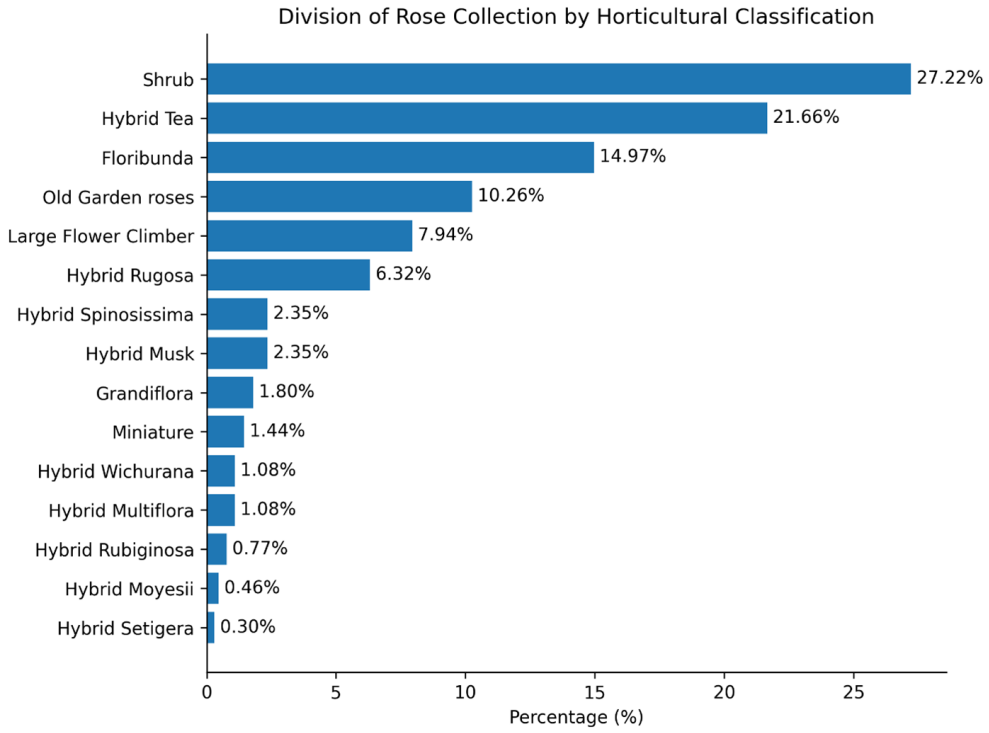


Figure 3. Division of cultivated varieties in the M.M. Gryshko NBG rose collection by horticultural classification

Cultivated varieties and heritage roses

Cultivated roses form the largest part of the collection and include a significant number of historical and heritage cultivars bred during the 19th and early 20th centuries. These accessions are maintained as clonal material and are of particular value for preserving ornamental traits, fragrance, and growth habits that are often absent in modern commercial cultivars.

The presence of old garden roses enables long-term comparative studies and supports the conservation of horticultural heritage. Many of these cultivars are maintained in a limited number of collections worldwide, emphasizing the importance of the M.M. Gryshko NBG as a conservation site.

The collection is divided per the Horticultural Classification of the American Rose Society (Young and Schorr, 2007) into 24 horticultural groups: 89.74% modern roses and 10.26% old garden roses (Centifolia, Hybrid Perpetual, Damask, Hybrid Gallica, Bourbon, Hybrid China, Noisette, Hybrid Foetida, Alba, Portland) (Figure 3).

Geographic origin of accessions

The geographic origin of accessions reflects the historical development of the collection and international exchange networks. A substantial proportion of cultivated material originates from Western and Central Europe, particularly Germany, France and the United Kingdom, regions that played a key role in the development of modern rose breeding since the late 19th century (Figure 4).

The initial core of the collection was established using planting material introduced from Germany in the mid-20th century. Subsequent enrichment was achieved through targeted acquisitions from botanical gardens, research institutions and breeders in Eastern Europe, North America, and Asia. This diverse geographic background contributes to the broad adaptive potential observed within the collection. The introduced cultivars originate from different breeding centres (see Figure 4).

Ukrainian cultivars

An important component of the collection consists of original Ukrainian rose cultivars developed through classical breeding programmes at the M.M. Gryshko NBG. These cultivars were selected for ornamental value, winter hardiness and suitability for landscape use under the climatic

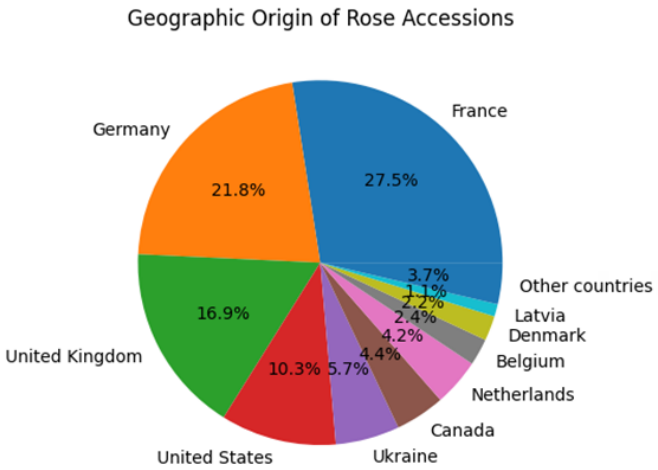


Figure 4. Geographic origin of rose cultivars in the M.M. Gryshko NBG collection

conditions of northern Ukraine (Figure 4).

To date, 12 Ukrainian cultivars (Table 2) have been officially registered, and all are conserved within the collection as reference material. Their inclusion ensures the preservation of national breeding achievements and provides a genetic basis for future breeding and research activities (State Register, 2025).

Conservation and management of the collection

The rose genetic resources of the M.M. Gryshko NBG are conserved primarily as a living field collection, which represents the most appropriate *ex situ* conservation strategy for clonally propagated ornamental plants. This approach ensures the preservation of cultivar identity, phenotypic stability and long-term availability of material for research, breeding and evaluation (Table 3).

To ensure long-term conservation and reduce the risk of genetic erosion, selected accessions of the rose collection are maintained as safety duplicates within different sectors of the M.M. Gryshko NBG, as well as in cooperating botanical institutions. The collection has a long tradition of exchanging plant material with botanical gardens, research institutions and breeding centres, primarily within Ukraine and, historically, with institutions in other European countries. Material exchange is carried out in compliance with national regulations and institutional policies and is mainly aimed at conservation, research and educational purposes. The proportion of accessions maintained as safety duplicates within the Garden or in other institutions is not constant, and depends on the conservation value of particular accessions, their propagation status and collaboration activities. Priority is given to valuable, rare, or vulnerable genotypes.

Table 2. Rose cultivars developed at the M.M. Gryshko NBG and their main characteristics

No.	Cultivar name	Passport data	Class	Main characteristics
1	Akvarel Rose Park	2013	Hybrid Tea	Height 90–100cm, flower colour: pink
2	Halaktyka	2021	Shrub	Height 200cm, flower colour: pink
3	Hratsiinyi Tanok	2011	Shrub	Height 130–150cm, flower colour: red
4	Karusel	2018	Shrub	Height up to 250cm, flower colour: pink
5	Khortytsia	2007	Shrub	Height up to 250cm, flower colour: yellow
6	Kraplia Sontsia	2025	Shrub	Height up to 200cm, flower colour: pink
7	Mushli	2021	Shrub	Height up to 200cm, flower colour: pink
8	Pochaina	2020	Shrub	Height up to 200cm, flower colour: pink
9	Solodkyi Son	2021	Shrub	Height up to 250cm, flower colour: pink
10	Spohady	2019	Floribunda	Height 130-150cm, flower colour: pink
11	Vintazh	2017	Shrub	Height up to 250cm, flower colour: yellow
12	Vrazhennia	2013	Shrub	Height up to 250cm, flower colour: crimson

Table 3. Structure of the *Rosa* genetic resources collection conserved *ex situ* at the M.M. Gryshko NBG rose collection

Component	No. of accessions	Conservation form	Characteristics and significance
Wild *Rosa* species	18	Living plants; seeds	Sources of adaptive traits (winter hardiness, drought tolerance, disease resistance); high conservation value
Non-Ukrainian cultivars (modern and old garden roses)	> 650	Living clonal plants	Cultivars of international origin representing major horticultural groups, including heritage germplasm
Ukrainian cultivars	12	Living clonal plants	National breeding products adapted to local environmental conditions
Total	> 668	Field *ex situ* collection with complementary seed storage	Largest *Rosa* genetic resources collection in Ukraine

Field *ex situ* conservation

All rose accessions are maintained under open-ground conditions in the collection plots of the M.M. Gryshko National Botanical Garden, covering an area of approximately 0.5ha.

The spatial arrangement of plants ensures adequate air circulation and access for maintenance operations, while minimizing disease pressure. Standard horticultural practices are applied to maintain plant health and longevity, while avoiding intensive management that could mask genotypic differences in stress tolerance.

The plants were grown on dark grey podzolized soils with a well-developed cultivated layer, enriched with compost applied in autumn. During the growing season, a complete mineral fertilizer (NPK 16–16–16; $NH_4H_2PO_4$ + NH_4NO_3 + KCl) was applied at 4-week intervals. Weeds are removed by hand weeding.

Irrigation was carried out every 14 days in the absence of rainfall. Less winter-hardy roses were protected using an air-dry shelter (covering with insulating materials) (Figure 2b), whereas more winter-hardy cultivars were overwintered by soil mounding (hilling-up). Frost damage is evaluated visually, based on the extent of injury to shoots and buds after winter.

Pruning practices varied among garden groups. Old garden roses were subjected to minimal pruning. Modern roses (Hybrid Tea and Floribunda) were pruned to approximately half of their natural, unpruned height. For climbing roses, long vigorous shoots were trained horizontally or obliquely, while lateral shoots were shortened to a well-developed bud.

Such conditions enable the long-term assessment of winter hardiness, frost damage, and general plant resilience, which are critical traits for the selection of cultivars suitable for landscaping in northern Ukraine. General plant resilience is assessed visually as an integrated trait, taking into account winter survival, the degree of recovery after frost damage, shoot regrowth, and overall plant vigour during the growing season.

Propagation and maintenance strategy

Cultivated varieties are conserved as clonal accessions and propagated vegetatively to maintain trueness-to-type. Propagation is carried out periodically to replace aging plants, restore damaged accessions, or increase the number of plants of particularly valuable genotypes.

Wild *Rosa* species are conserved both as living plants and, where possible, through seed propagation. This dual strategy allows the preservation of genetic diversity within species while ensuring the availability of material for regeneration and research.

Accession identity is maintained through careful labelling, long-term curatorial records, and expert morphological assessment based on stable diagnostic traits, including plant habit, flower morphology and flowering behaviour.

Seed collection and complementary conservation

Living plants of wild species are maintained in the field, and seed samples are preserved in the seed laboratory of the M.M. Gryshko NBG, providing an additional level of conservation.

From the plants cultivated in the collection, seed material is obtained, then cleaned and dried at room temperature. Later, part of the seeds is used through Index Seminum for international scientific exchange with botanical institutions around the world. These seeds are stored in paper bags in an amount of 200–400g at room temperature and are updated annually. Such storage conditions are quite sufficient to preserve the seeds for 1–2 years without significant loss of viability. Different types of rose hips retain germination even without special treatment from 20% (*R. canina* L.) to 40% (*R. rugosa* Thunb.). The rest of the seeds, which are used for sowing in breeding work, as well as those obtained under the scientific exchange system, undergo stratification in a refrigerator at a temperature of approximately 5°C, which increases the germination of different types of rose hips to 60–80%.

In addition, the seed laboratory of the M.M. Gryshko NBG keeps a carpological collection that demonstrates the diversity of morphological types of fruits and seeds of the world's flora. The genus *Rosa* is represented by 160 specimens, covering 73 species of rose hips. In the carpological collection, dried seeds weighing 20–40g are stored in glass stoppered test tubes in cabinets at room temperature (Figure 2f). Such specimens can be stored for many years. Due to the loss of germination, they are not intended for sowing, but are used for anatomical or morphological studies in resolving controversial phylogenetic and systematic issues, and also serve as didactic material for garden employees, students and postgraduates studying the carpology of various systematic groups of plants. In the future, it is planned to create a full-fledged genetic seedbank on the basis of the seed laboratory with the involvement of special equipment and compliance with international requirements for the storage of genetic resources.

Perspectives for long-term conservation

The rose collection of the M.M. Gryshko NBG (Table 3) is officially recognized as an object of national scientific heritage, which provides an institutional framework for its long-term conservation (Rakhmetov *et al*, 2019). Future priorities include the continued enrichment of the collection with underrepresented horticultural groups and wild species, improvement of documentation and data standardization, and integration into international networks for ornamental plant genetic resources, which would enable the exchange of plant material and information, harmonization of data standards, participation in collaborative research, and the development of coordinated conservation strategies.

Under current climate change scenarios, the collection also serves as a testing ground for identifying genotypes with enhanced tolerance to abiotic stresses, reinforcing its strategic importance for sustainable landscaping and breeding.

Documentation and data availability

Comprehensive documentation constitutes a core component of the *ex situ* conservation of rose genetic resources at the M.M. Gryshko NBG. All accessions maintained in the collection are supported by curatorial records that ensure traceability, long-term management, and availability of data for research and conservation purposes.

The internal documentation system of the M.M. Gryshko NBG is primarily based on curated archival records maintained as structured paper files. These records include passport data, inventory information and documentation of

collection history. Selected data are additionally maintained in digital form using standard spreadsheet tools to facilitate data handling and updating, with ongoing efforts toward further digitalization and data standardization.

In addition to textual records, the collection is supported by a photographic archive documenting key morphological traits, phenological stages and general plant habit. These images are used for verification of accession identity and for comparative analysis.

Passport data include basic information on accession identity (scientific name, cultivar denomination, origin, source, and year of introduction), while detailed morphological and ornamental traits (including floral characteristics) are recorded separately within the characterization and evaluation (C&E) dataset.

Documentation data are available upon request for scientific, educational and conservation purposes. At present, the data are not publicly accessible online; however, selected information is disseminated through scientific publications and institutional reports, including studies on rose collections and their morphological and biological traits (Rubtsova *et al*, 2021; Rubtsova *et al*, 2022; Rubtsova *et al*, 2023; Rubtsova *et al*, 2025a; Rubtsova *et al*, 2025b).

Passport data

Each accession is accompanied by passport data, including the scientific name, cultivar denomination, horticultural group, geographic origin, source of acquisition and year of introduction into the collection. For cultivated varieties, additional information on breeder, country of origin and breeding period is recorded where available.

Passport data are maintained in the internal documentation system of the M.M. Gryshko NBG and are regularly updated during collection inventory and regeneration activities. This information allows the reconstruction of accession history and supports the assessment of collection completeness and representativeness.

Accession identification and quality control

Accession identity is ensured through expert morphological verification, supported by historical records and comparison with reference descriptions. Periodic assessments are carried out to detect possible mislabelling, somatic mutations, or loss of clonal integrity, which is particularly relevant for long-maintained ornamental cultivars (Rubtsova and Chyzhankova, 2026).

In cases of uncertainty, accessions are flagged for further evaluation, and propagation material is taken from verified plants to maintain collection accuracy.

This quality control is essential for long-term genebank-type conservation of clonally propagated ornamentals.

Data accessibility and exchange

Documentation data are available upon request for scientific, educational and conservation purposes, in accordance with institutional regulations of the M.M. Gryshko NBG. Exchange of plant material and associated data is conducted through bilateral agreements with botanical gardens and research institutions.

M.M. Gryshko NBG follows internationally accepted principles for the exchange and use of plant genetic resources, ensuring transparency and responsible use of conserved material. Future efforts are aimed at improving data standardization and facilitating integration into international information systems for plant genetic resources.

Use of the collection

The rose genetic collection conserved *ex situ* at the M.M. Gryshko NBG serves multiple functions, including scientific research, breeding, education and public outreach, thereby reinforcing its role as a national reference collection of ornamental plant genetic resources (Rubtsova and Chizhankova, 2017; Gordienko *et al*, 2021).

Research and long-term observations

Characterization and evaluation of rose accessions are conducted through long-term field observations under uniform environmental conditions. C&E data include morphological descriptors (plant habit, flower form, colour, flowering period) (Rubtsova *et al*, 2022; Rubtsova *et al*, 2023), phenological traits (Boiko at al, 2015; Pokhylchenko at al, 2024), and indicators of ornamental value (e.g. flowering abundance, duration, and overall decorative effect in plantings), as applied in urban landscaping studies (Kolesnichenko *et al*, 2020).

The oldest data used in this study date back to 1958 and are derived from the monograph by L.P. Lempitskyi (1958), which represents one of the earliest comprehensive sources on rose cultivation in Ukraine.

Particular attention is given to adaptive traits of practical relevance, such as winter hardiness, drought tolerance and resistance to major fungal diseases. These traits are assessed over multiple growing seasons, enabling the identification of stable phenotypic patterns and genotypes with enhanced adaptive potential.

C&E data are used to support scientific publications, selection of material for breeding, and recommendations for landscape use. Part of these data has already been published, including studies on frost tolerance and morphological traits of *Rosa rugosa* cultivars (Rubtsova and Chyzhankova, 2017), adaptation of Scots roses in Northern Ukraine (Rubtsova *et al*, 2021), adaptation of East Asia roses (Rubtsova at al, 2025b) and resistance to powdery mildew (Rubtsova *et al*, 2025a). While full datasets are not currently available in international online databases, summarized information is disseminated through peer-reviewed publications and institutional reports.

The collection provides a stable experimental base for long-term studies on rose biology, phenotypic variability, and adaptation to environmental conditions. Uniform cultivation under open-ground conditions allows comparative assessment of accessions with different genetic backgrounds and horticultural groups.

Long-term observations have supported research on somatic mutations (Rubtsova *et al*, 2026), phenotypic stability, flowering behaviour, and adaptive traits such as winter hardiness and drought tolerance. Results obtained using material from the collection have been disseminated through peer-reviewed scientific publications, contributing to the understanding of rose diversity and adaptation in temperate climates (Rubtsova *et al*, 2025a).

Breeding and selection

The collection constitutes an important genetic reservoir for rose breeding programmes conducted at the M.M. Gryshko NBG. Wild species and selected cultivars are used as parental material in classical breeding schemes aimed at developing varieties adapted to local climatic conditions.

Particular emphasis is placed on the selection of genotypes combining high ornamental value with increased tolerance to abiotic stresses. Ukrainian cultivars conserved in the collection represent tangible outputs of this work and serve as reference material for further breeding and evaluation.

Education and professional training

In cooperation with the National University of Life and Environmental Sciences of Ukraine, the rose garden contributes to university-level teaching, including courses on phenological observations, rose breeding (Figure 2g) and the use of roses in landscape design. It also organizes thematic exhibitions, such as the 'Shakespeare's Roses' display dedicated to the Year of Shakespeare.

The collection also serves as a platform for professional training of specialists in ornamental horticulture and for the dissemination of knowledge on plant genetic resource conservation. In addition, the rose garden is used as a venue for cultural events, including open-air concerts (Figure 2h), which enhance public engagement and promote awareness of plant diversity and the aesthetic value of cultivated roses.

Public outreach and cultural value

In addition to its scientific role, the rose collection has significant cultural and educational value for the general public. It is an integral part of the M.M. Gryshko NBG's exhibition infrastructure and attracts a wide audience during the flowering season.

By maintaining historical and heritage cultivars alongside modern varieties, the collection contributes to the preservation of horticultural traditions and raises public awareness of the importance of conserving plant genetic diversity.

Conclusions

The rose genetic collection conserved *ex situ* at the M.M. Gryshko NBG represents the largest and most diverse collection of *Rosa* genetic resources in Ukraine. Apart from the M.M. Gryshko NBG, several other institutions maintain significant rose collections. Rose collections in Ukraine are maintained in several leading botanical institutions. Significant rose garden are also established at the National Dendrological Park Sofiyivka and the Dendrological Park Alexandria of the National Academy of Sciences of Ukraine. Important collections of *Rosa* spp. are further held in university botanical gardens, including the O.V. Fomin Botanical Garden of Taras Shevchenko National University of Kyiv, the Botanical Garden of Odesa I.I. Mechnikov National University, and the Botanical Garden of Yuriy Fedkovych Chernivtsi National University, Botanical Garden of Oles Honchar Dnipro National University. A historically important centre of rose cultivation and introduction is the Nikita Botanical Garden – National Scientific Center (Crimea), which has long contributed to the development and study of ornamental roses in the region.

Maintained primarily as a living field collection, it ensures the long-term preservation of cultivated varieties, wild species, and national breeding achievements.

Systematic conservation, documentation, and long-term evaluation under open-ground conditions provide valuable data on adaptive traits of roses relevant to temperate climates. The collection supports research, breeding, education and public outreach, thereby fulfilling key objectives of *ex situ* conservation of ornamental plant genetic resources.

By documenting the structure, management, and use of this collection in a standardized format, the present article contributes to the visibility of Ukrainian ornamental plant genetic resources and their integration into the international framework of plant genetic resource conservation.

Author contributions

Olena Rubtsova: conceptualization and ideation. Mykola Shumyk: writing, review, and editing. Natalia Chuvikina: historical background. Tetyana Vakulenko: writing, data analysis. Valentina Chizhankova: data collection.

Conflict of interest statement

The authors declare that they have no conflicts of interest.

References

Ahr, H., and Lang, I. (1991). Zeittafel des Rosarium Sangerhausen. Rosen-jahrbuch 1990, ed. Schoser G. (Verlag GmbH & Co.), 147–159.

Boiko, R.V., Shcherbakova, O.F., Rubtsova, O.L., & Chyzhankova, V.I. (2015). Methodological recommendations for phenological observations of recurrently flowering roses (Ukraine : Veles Publishing), 52 p.

Cherevchenko T.M., Chuvikina N.V. (2003). The role of Danilo Fedorovitch Likhvar in the formation of the Central Republican Botanical Gardens of Academy of Sciences of Ukraine (on the 100-th anniversary of birth). *Plant Introduction* 1-2, 166–172. url: https://www.plantintroduction.org/index.php/pi/article/view/1143/1099

Gordienko D.S., Rubtsova O.L., Buidina T.O., Chizhankova V.I., Rozhok O.F., Sokolova O.A. (2021). Drought resistance of English roses varieties in the conditions of Right-bank Forest-Steppe of Ukraine. Plant Varieties Studying and Protection 17, 60-65. https://doi.org/10.21498/2518-1017.17.1.2021.228212

Gryshko, M.M. (1949). Objectives and Directions of the Work of the Botanical Garden of the Academy of Sciences of the Ukrainian SSR. In *Proceedings of the Botanical Garden of the Academy of Sciences of the Ukrainian SSR*. Kyiv, 1, 3–21.

Collective (2023). Guide des Roseraies en Europe (Paris, France: Société Nationale d'Horticulture de France), 244 p.

Kolesnychenko, O.V., Rubtsova, O.L., Shumyk, M.I., Hrysiuk, S.M., Pikovskyi, M.Y., Shvets, I.V., Morozko, A.P., & Chyzhankova, V.I. (2020). Roses in urban plantings of Kyiv (Ukraine: Yamchynskyi Publishing), 267 p.

Krüssmann, G. (1981). The Complete Book of Roses (USA, Portland: Timber Press), 436 p.

Lempitskyi, L.P. (1958). Kultura roz u vidkrytomu hrunti [Rose cultivation in open ground]. (Ukraine: Publishing House of the Academy of Sciences of the Ukrainian SSR), 124 pp.

Lempitsky L.P., Halytska A.F. (1968) Introduction of roses

in the Ukrainian SSR. Introduction and acclimatization of plants in Ukraine 3, 39–45.

Pokhylchenko, O., Gorelov, O., Dovhalyuk, N., Kozubenko, T., Kruglyak, Y. and Paraschuk, O. (2024). Phenological observations in dendrological collections: motivation for the use of the universal coding "BBCH scale". *Journal of Native and Alien Plant Studies* 20, pp. 176–197. https://doi.org/10.37555/2707-3114.20.2024.318688

Rakhmetov D.B., Zaimenko N.V., Gaponenko M.B., Buyun L.I., Rubtsova O.L. (2019). Scientific objects of the M.M. Gryshko NBG of the National Academy of Sciences of Ukraine, firming the national heritage (Ukraine: Palivoda A.V. Publishing), 224 p.

Red Data Book of Ukraine. Plant Kingdom (2009). Ed. by Ya.P. Didukh. Kyiv: Globalconsulting. Available at: https://redbook-ua.org/

Rubtsova O., Chizhankova V. (2017). Morphological characteristics and frost tolerance of cultivars derived from the *Rosa rugosa* Thunb. *Journal of Applied Horticulture* 19 (1). 67–69. doi: http://dx.doi.org/10.37855/jah.2017.v19i01.11

Rubtsova O., Chyzhankova V. (2026). Somatic mutations in introduced garden rose cultivars in the collection of the M.M. Gryshko National Botanical Garden. *Plant Introduction* 109, 55-64. https://doi.org/10.46341/PI2025031

Rubtsova, O., Kolesnichenko, O., Shumyk, M., Shynder, O., Chyzhankova, V., Dzyba, A. and Hrysiuk, S. (2021). Scots roses in Northern Ukraine. *Horticultural Science* 48(3), 144–148. https://doi.org/10.17221/90/2020-HORTSCI

Rubtsova, O.L., Vakulenko, T.B. and Chyzhankova, V.I. (2022) Morphological features of nectaries of some species of the genus *Rosa* (Rosaceae). *Ukrainian Botanical Journal* 79(2), 103–113. https://doi.org/10.15407/ukrbotj79.02.103

Rubtsova, O.L., Vakulenko, T.B., Chyzhankova, V.I. (2023) Teratogenesis of generative organs of cultivars of *Rosa* (Rosaceae) in the collection of the M.M. Gryshko National Botanical Garden, National Academy of Sciences of Ukraine. *Ukrainian Botanical Journal* 80(4), 364–374. https://doi.org/10.15407/ukrbotj80.04.364

Rubtsova O., Pikovskyi M., Myroshnychenko D., Chyzhankova V. (2025a). Evaluation of powdery mildew (*Podosphaera pannosa* (Wallr.) de Bary) in roses. *Biological Systems: Theory and Innovation* 16, 1, p.22–32. https://doi.org/10.31548/biologiya/1.2025.22

Rubtsova O., Shumyk M., Shynder O., Chyzhankova V. (2025b). Roses of East Asia in Ukraine (Ukraine: Feniks Publishing), 160 p. https://www.researchgate.net/publication/400214857_SIPSINI_SHIDNOI_AZII_V_UKRAINI_Monografia

Scientific Archive of the M.M. Gryshko National Botanical Garden, National Academy of Sciences of Ukraine (1946), Reports on scientific expeditions of the Botanical Garden staff (missions to Germany). Fond 2, file 1, 103 pp.

State Register of Plant Varieties Suitable for Distribution in Ukraine (2025). Available at: https://data.gov.ua/dataset/eabd0bd2-2dc6-47e2-b748-9bd254da4956

Young, M. A. and Schorr, P. (2007). Modern Roses 12: The comprehensive list of roses in cultivation or of historical or botanical importance (USA, Shreveport, Louisiana: American Rose Society), 576 p.

Genetic Resources

ORIGINAL ARTICLE

Genetic Resources (2026), 7 (13), 153–168
DOI: 10.46265/genresj.ROKV2181
https://www.genresj.org
ISSN: 2708-3764

Characterization of Iranian rice genetic resources for key grain quality traits

Mostafa Modarresi

Rice Research Institute of Iran, Agricultural Research, Education and Extension Organization (AREEO), Rasht, Islamic Republic of Iran

Abstract: Grain quality traits are crucial for rice (*Oryza sativa* L.) breeding, as they significantly influence consumer preferences and market value. This study evaluated 48 diverse rice genotypes, including a strong representation of Iranian landraces that constitute unique and valuable genetic resources at risk of erosion, as well as some improved varieties and imported lines, for seven grain quality traits: amylose content, gelatinization temperature, grain length, grain width, cooked grain length, grain shape, and grain elongation during the 2023–2024 growing seasons. The studied rice genotypes showed substantial variation in quality-related traits. All traits demonstrated significant genotypic variation, with high broad-sense heritability ($H^2 \geq 0.968$) and notable genetic advance (up to 49.37% for gelatinization temperature), indicating strong potential for genetic improvement. Correlation analyses showed strong positive relationships among grain length, grain shape, and cooked grain length, but negative associations with grain width and grain elongation, indicating breeding trade-offs. Path analysis highlighted grain length as a primary driver of cooked grain length and identified gelatinization temperature and grain shape as key influences on amylose content. Hierarchical clustering and principal component analysis identified four genotype clusters, with standout performers like Gohar (excelling in grain length) and Gharib Siah Reyhani (high elongation) suitable for breeding programmes. These results not only provide a robust framework for multi-trait selection and the development of high-quality varieties tailored to Iranian and global markets, but also underscore the conservation value and uniqueness of Iranian landraces as representative genetic resources for future rice improvement and food security.

Keywords: *Oryza sativa*, grain quality, genotypic variation, heritability, path analysis, breeding

Citation: Modarresi, M. (2026) “Characterization of Iranian rice genetic resources for key grain quality traits”, *Genetic Resources*, 7(13), pp. 153–168. doi: 10.46265/genresj.ROKV2181.

Introduction

Rice (*Oryza sativa* L.) is a staple food for over half of the global population, providing a primary source of calories and nutrition, particularly in Asia, Africa and Latin America (Birla *et al*, 2017). The quality of rice grains, encompassing physical, chemical and cooking properties, significantly influences consumer preferences, market value and food security. Key grain quality traits, such as amylose content, gelatinization temperature, grain length, grain width, cooked grain length, grain shape and grain elongation, determine the sensory and culinary attributes of rice, including texture, appearance and cooking behaviour (Hori and Sun, 2022; Modarresi, 2023). For instance, amylose content affects the stickiness and firmness of cooked rice. In contrast, cooked grain length and grain shape contribute to the visual appeal and palatability preferred in premium rice varieties. These traits are under complex genetic control, influenced by both genotypic and environmental factors, making their improvement a priority in rice breeding programmes (Sultana *et al*, 2022).

Breeding for enhanced grain quality requires a thorough understanding of genotypic variation, trait interrelationships, and genetic parameters such as heritability and genetic advance. High heritability and substantial genetic variation indicate traits amenable to selection, while correlation and path analyses reveal how traits interact to influence overall quality (Kumar *et al*, 2010; Thuy *et al*, 2023). For example, grain length and cooked grain length are often positively correlated, suggesting that selecting for longer

* Corresponding author: Mostafa Modarresi (mstmodarresi@gmail.com)

Received: 21.10.2025 | Accepted: 05.03.2026 | Published online: 06.05.2026

grains may enhance cooked grain appearance (Malik *et al*, 2022). However, trade-offs, such as the negative association between grain width and grain shape, necessitate careful consideration in breeding strategies. Multivariate techniques, including principal component analysis (PCA) and hierarchical clustering, further aid in identifying genotypes with desirable trait combinations and understanding the underlying structure of trait variation (Talekar *et al*, 2022). Despite advances in rice breeding, comprehensive studies integrating these approaches to evaluate diverse germplasm, particularly in regions like Iran with rich rice genetic diversity, remain limited.

Iran is a significant rice-producing country, with a wide array of local landraces and improved cultivars adapted to diverse agroecological conditions (Modarresi, 2023; Modarresi *et al*, 2024). These local landraces, shaped by centuries of farmer selection, represent a uniquely rich component of Iran's rice genetic resources, harbouring trait combinations – particularly for grain quality – that are often absent or diluted in modern improved varieties and imported lines. Their representativeness of traditional Iranian rice diversity, coupled with their adaptation to local environments, confers high conservation value, especially as many are threatened by genetic erosion due to the widespread adoption of high-yielding modern cultivars. These genotypes offer a valuable resource for improving grain quality traits, yet their genetic potential and trait associations are underexplored.

Previous studies in Iran, such as Abarshahr *et al* (2011), Yadi *et al* (2021), Safiedin Ardebili *et al* (2024), have focused primarily on yield-related traits, morphometric diversity of individual cultivars, or limited quality parameters examined in isolation, often without integrating genetic parameters, trait interrelationships and multivariate analyses. To our knowledge, no prior Iranian study has provided a comprehensive, multi-trait evaluation that simultaneously characterizes genotypic variation, estimates broad-sense heritability and genetic advance, examines trait associations through correlation and path coefficient analyses, and applies principal component analysis and hierarchical clustering to identify superior genotype combinations across the seven key physical and chemical grain quality traits most relevant to cooking, eating and market value.

This study aimed to evaluate 48 rice genotypes for seven grain quality traits that align closely with preferred cooking and eating characteristics to inform breeding for improved grain quality. Specific objectives were to: (1) assess phenotypic variation and genotypic differences in amylose content, gelatinization temperature, grain length, grain width, cooked grain length, grain shape and grain elongation; (2) estimate genetic parameters, including heritability and genetic advance, to identify traits with high breeding potential; (3) examine trait interrelationships using correlation, PCA and path analysis to understand their contributions to grain quality; and (4) identify superior genotypes with desirable trait combinations through clustering and multivariate analyses. By integrating these approaches on a highly representative Iranian germplasm panel, this study provides a novel framework for multi-trait selection and conservation-informed breeding.

Materials and methods

Plant material and experimental design

The rice germplasm was obtained from the Rice Research Institute of Iran (RRII), located in Rasht, Gilan Province. It consisted of 48 genotypes, including a diverse set of local landraces, RRII-improved/registered cultivars, RRII-improved/breeding lines, and imported genotypes (Table 1), and was evaluated for grain quality traits. This composition provides a highly representative sample of Iranian rice genetic diversity while incorporating unique landrace material of considerable conservation value. The genotypes included Deylamani, Hashemi, Shiroudi, Binam, Gohar and 43 others, selected to represent a broad range of genetic diversity in Iranian and international rice germplasm. All 48 genotypes are maintained as pure lines at RRII. Seeds for the experiments were obtained from uniform multiplication plots grown under standard conditions at RRII in the season preceding the 2023 trial to ensure genetic purity, viability and consistency across genotypes. The experiments were conducted in a randomized complete block design (RCBD) with four replications at RRII, Rasht, Iran, during the 2023 and 2024 growing seasons. Each genotype was planted in a 2m² plot, with standard agronomic practices applied uniformly across all plots to minimize environmental variation. The four replications ensured robust estimation of genotypic effects while accounting for field heterogeneity.

Trait measurements

Seven grain quality traits were measured on harvested grain samples from each plot (4 replications per year × 2 years), yielding 384 plot-level observations in total (48 genotypes × 4 replications × 2 years). The traits included amylose content (%), determined using the colourimetric method with iodine-potassium iodide solution, following the protocol of Juliano (1971). Absorbance was measured at 620nm using a spectrophotometer, and amylose content was calculated based on a standard curve. Gelatinization temperature was assessed using the alkali spreading test (Little, 1958), where grains were soaked in 1.7% potassium hydroxide solution for 23 hours at 30°C. Spreading was scored on a 1–7 scale, with higher scores indicating lower gelatinization temperature. Although this alkali spreading value (ASV) is technically an ordinal score, it is routinely treated as a quasi-continuous variable in rice grain quality and breeding studies for parametric statistical analyses. This approach is widely accepted due to the scale's progressive correspondence with actual gelatinization temperature and its consistent use in ANOVA, correlation, heritability, and multivariate analyses in the literature (e.g. Little, 1958; Juliano, 1971; Pang *et al*, 2016; Xu *et al*, 2024). Grain length (mm) was measured as the average length of ten randomly selected whole milled grains per plot. Grain width (mm) was measured as the average width of the same ten grains, perpendicular to the length. Cooked grain length (mm) was determined by measuring ten cooked grains selected from a sample of 20–100 whole, milled rice grains, presoaked in distilled water for 30 minutes at ambient temperature, cooked in a vigorously boiling water bath at 100°C for ten minutes, cooled in a water bath at ambient temperature, and blotted with blotting paper (Juliano and Perez, 1984). The 10-grain average provided the plot-level value. Grain shape was calculated as the ratio of

Table 1. Origin and year of release of genotypes used in this study. *Cy, Dcl, and Line7 are historical codes for imported rice genotypes from international collections, developed more than 40 years ago; detailed full names are unavailable due to limited archival records.

Genotype	Origin	Year of release	Genotype	Origin	Year of release
Deylamani	Iranian landrace, Mazandaran province	-	Fuji Minori	Imported/Japan/Registered cultivar	1971
Hashemi	Iranian landrace, Guilan province	-	Sepidroud	RRII-improved/Registered cultivar	1987
Shiroudi	RRII-improved/Registered cultivar	2008	Rash	IRRI-improved/Registered cultivar	2017
Binam	Iranian landrace, Guilan province	-	Koohsar	IRRI-improved/Registered cultivar	2011
Gohar	RRII-improved/Registered cultivar	2016	Gharib Siah Reyhani	Iranian landrace/Guilan province	-
Khazar	RRII-improved/Registered cultivar	1983	Neda	RRII-improved/Registered cultivar	1998
Hassani	Iranian landrace, Guilan province	-	Anam	RRII-improved/Registered cultivar	2018
Saleh	RRII-improved/Registered cultivar	2002	Zenith	Imported/United States	-
Abji Boji	Iranian landrace, Guilan province	-	Dorfak	RRII-improved/Registered cultivar	2001
Anbarboo Ilam	Iranian landrace, Ilam province	-	Keshvari	IRRI-improved/Registered cultivar	2011
Champa Boodar	Iranian landrace, Guilan province	-	Fajr	IRRI-improved/Registered cultivar	2001
Gharib	Iranian landrace, Guilan province	-	Bejar	RRII-improved/Registered cultivar	1993
Salari	Iranian landrace, Mazandaran province	-	Line7*	Imported/IRRI	-
Ali Kazemi	Iranian landrace, Guilan province	-	Nemat	RRII-improved/Registered cultivar	1995
Mutant Gohar	RRII-improved/Breeding line	-	Ahlami Tarom	Iranian landrace, Mazandaran Province	-
Tisa	RRII-improved/Registered cultivar	2018	Dom Siah	Iranian landrace, Guilan province	-
Hassan Saraei	Iranian landrace, Guilan province	-	Shahpasand	Iranian landrace, Mazandaran Province	-
TH1	RRII-improved/Breeding line	-	Ghodsi	RRII-improved/Breeding line	-
Tetep	Imported/Vietnam	-	Sange Tarom	Iranian landrace, Mazandaran Province	-
Tarom Mahalli	Iranian landrace, Mazandaran Province	-	Dom Zard	Iranian landrace, Guilan province	-
Anbarboo	Iranian landrace, Khuzestan province	-	Dcl*	Imported/Egypt	-
Kadous	IRRI-improved/Registered cultivar	2003	Cy*	Imported/Egypt	-
Mohammadi Chaparsar	Iranian landrace, Mazandaran Province	-	IR64	Imported/IRRI	-
Dom Sefid	Iranian landrace, Guilan province	-	IR36	Imported/IRRI	-

grain length to grain width, reflecting grain slenderness. The grain shapes were categorized into three types based on their length-to-width ratio, displayed from top to bottom as follows: slender (ratio > 3), medium (ratio 2–3), and bold (ratio < 2) (Zhao *et al*, 2022a). Grain elongation calculated as the ratio of cooked grain length to uncooked grain length, indicating grain expansion during cooking. All measurable traits were based on plot-level averages (of ten grains where specified) or bulk samples. These plot-level values were used directly in descriptive statistics and ANOVA. Genotype means (averaged across the eight plot-level values per genotype) were used for correlation, hierarchical clustering, principal component analysis and path analysis. All measurements were conducted in a rice quality laboratory environment to ensure consistency.

Statistical analyses

Data were analyzed using R version 4.3.1 (R-Core-Team, 2023) for descriptive statistics, analysis of variance, correlation analysis, hierarchical clustering and principal component analysis, using packages including dplyr, agricolae, corrplot, Hmisc, ComplexHeatmap, dendextend, tidyverse, ggfortify, and factoextra. Descriptive statistics – mean, standard error, minimum, maximum, and range – were calculated for each trait across all observations using the dplyr package to summarize trait variability.

Analysis of Variance (ANOVA): To assess the validity of combining data across years, a preliminary full ANOVA model including the genotype × year (G × Y) interaction term was fitted for each trait using the aov function. The G × Y interaction was non-significant for all seven traits (F-values ranging from 0.91 (grain elongation) to 1.28 (amylose content); df = 47, $p > 0.10$). Homogeneity of error variances between years was confirmed using Levene's test (leveneTest function in the car package; $p > 0.05$ for all traits). Year was treated as a fixed effect. These results justified pooling data from the 2023 and 2024 seasons for the final combined analysis of variance (ANOVA), performed using the aov function with genotype (48 levels) as the main factor, year (2 levels) as a fixed effect, and replication (4 levels) as a blocking factor. Variance components for estimation of genetic parameters (including broad-sense heritability) were derived from this pooled ANOVA.

A combined ANOVA was performed across two growing seasons using the following linear model:

$$Y_{ijk} = \mu + G_i + Y_j + G \times Y_{ij} + R_k(Y_j) + \varepsilon_{ijk}$$

where is the observed value of a trait, is the overall mean, is the effect of the genotype, is the effect of the year, is the G × Y interaction, is the effect of the replication nested within year, and is the residual error.

Mean performance: Mean trait values were computed for each genotype by averaging across the four replications using dplyr, providing genotypic performance profiles. Genotypic means were compared using Tukey's Honest Significant Difference (HSD) test at a 5% significance level.

Genetic Parameters: Variance components were estimated from ANOVA results. Genetic variance ($\sigma^2 g$) was calculated as (MSg - MSe) / r, where MSg is the mean square for genotype, MSe is the mean square error, and r is the number of replications (4). Environmental variance ($\sigma^2 e$) was equated to MSe, and phenotypic variance ($\sigma^2 p$) was $\sigma^2 g + \sigma^2 e$. Broad-sense heritability (H^2) was computed as $\sigma^2 g / \sigma^2 p$. Genotypic and phenotypic coefficients of variation (GCV, PCV) were calculated as $(\sqrt{\sigma^2 g}/\text{mean}) \times 100$ and $(\sqrt{\sigma^2 p}/\text{mean}) \times 100$, respectively. Genetic advance (GA) was estimated as $GA = k \times \sqrt{\sigma^2 g} \times H^2$, where $k = 2.06$ (5% selection intensity), and expressed as a percentage of the mean.

Correlation analysis: Pearson correlation coefficients among the seven traits were calculated using the rcorr function in the Hmisc package, with significance tested at $p < 0.05$. A correlation heatmap was generated using corrplot to visualize relationships.

Hierarchical clustering and heatmap: Mean trait values per genotype were standardized (z-scores) and subjected to hierarchical clustering using Euclidean distance and Ward's method (ward.D2) in the ComplexHeatmap package. A heatmap was produced to visualize trait patterns across genotypes.

Principal component analysis (PCA): PCA was performed on standardized mean trait values using the prcomp function to identify major axes of variation. Biplots of genotypes and traits on the first two principal components (PC1, PC2) were generated using ggfortify and factoextra to visualize genotypic and trait contributions.

Path analysis (structural equation modelling) was conducted separately using Python, employing the semopy package, which is specifically designed for structural equation modelling and provides standardized path coefficients and widely accepted goodness-of-fit indices. The use of Python for path analysis was independent of the R-based analyses and allowed for flexible specification and evaluation of alternative causal models. Path analysis was conducted to disentangle direct and indirect relationships among grain quality traits based on an explicitly defined biological and physicochemical framework. Two separate path models were specified, with cooked grain length and amylose content selected as focal (dependent) traits in independent analyses. These traits were chosen because they represent integrative endpoints of rice cooking and eating quality rather than primary morphological descriptors. In the first model, cooked grain length was treated as a key post-cooking physical outcome influenced by pre-cooking grain morphology (grain length and grain width) and starch-related properties. Grain length and grain width were specified as upstream traits because they are determined during grain development and physically constrain the extent of kernel expansion during cooking. Grain shape (length-to-width ratio) was modelled as an intermediate composite trait influenced by grain length and width, consistent with its mathematical definition and biological interpretation as a descriptor of kernel slenderness. Grain elongation was modelled as a downstream response variable influenced by cooked grain length and grain shape. This directionality reflects the biological and operational definition of grain elongation as the relative increase in kernel length after cooking (i.e. cooked grain length relative to uncooked grain length). Thus, cooked grain length is a necessary antecedent of elongation, and elongation represents a secondary, derived cooking response rather than an independent primary trait. In the second model, amylose content was treated as the focal chemical trait, reflecting its central role in determining rice texture and cooking behaviour. Gelatinization temperature was specified as an upstream physicochemical trait influencing amylose behaviour during cooking, while grain shape was included as a morphological correlate reflecting known associations between kernel slenderness and starch composition. Covariances among biologically related traits (e.g. amylose content and gelatinization temperature; grain length and grain width) were specified where appropriate. All path models were hypothesis-driven and grounded in established rice grain development, starch chemistry and cooking-quality literature. Standardized path coefficients were estimated using structural equation modelling, and model fit was evaluated using comparative fit index (CFI ≥ 0.90) and root mean square error of approximation (RMSEA ≤ 0.08). The Python script used for this analysis is provided as Supplemental File 1.

Results

Descriptive statistics of rice grain traits

The seven grain quality traits exhibited considerable variation across the 48 rice genotypes evaluated over four replications per year across two years (Table 2). Amylose content ranged from 15.97% to 28.68% (mean = 22.88%), indicating diverse starch properties. Gelatinization temperature varied from 3.13 to 7 (mean = 4.95), reflecting differences in cooking characteristics. Grain length and cooked grain length showed means of 7.21mm and 11.56mm, respectively, with ranges of 5.01–8.33mm and 8.47–14.37mm. Grain width averaged 2.14mm, ranging from 1.54 to 3.06mm. Grain shape, a ratio of length to width, had a mean of 3.42 (range = 1.70–5.48), while grain elongation, indicating cooking expansion, averaged 1.61 (range = 1.36–2.09). These results highlight substantial phenotypic diversity among genotypes, critical for breeding programmes targeting grain quality.

Analysis of variance (ANOVA)

Prior to pooling, a preliminary full model confirmed non-significant G × Y interaction for all traits ($F = 0.91–1.28$, $df = 47$, $p > 0.10$), together with homogeneous error variances across years (Levene's test, $p > 0.05$). This justified combining the data across years for the final ANOVA. The effect of year (comparing the 2023 and 2024 growing seasons) was not statistically significant according to the ANOVA results, indicating that the measured traits remained consistent between the two years (Table 3). Two-way ANOVA revealed highly significant genotypic effects ($p < 0.001$) for all seven traits, underscoring substantial genetic variation. The calculated F-values for the genotype effect, derived from the mean squares presented in Table 3, ranged from 21.47 for grain elongation to 266.79 for gelatinization temperature, confirming strong differentiation among the 48 genotypes. In contrast, replication effects were non-significant ($p > 0.05$) for all traits, with corresponding F-values below 0.99, suggesting minimal environmental variation due to experimental blocks. These findings confirm that observed trait differences are primarily driven by genetic factors, facilitating effective selection for desired grain characteristics.

Mean performance of genotypes

Mean comparisons among the studied genotypes were performed using Tukey's honest significant difference (HSD) test at a 5% significance level to determine significant differences among treatments. The grouping letters shown in Table 4 represent the statistical comparison results, where means followed by the same letter are not significantly different from each other. This analysis provided a clear distinction of genotypic performance for each evaluated trait. For instance, Saleh exhibited the highest amylose content (27.1%) and grain shape (5.3), while Gohar had the longest grain (8.32mm) and cooked grain length (14.08mm) (Table 4). Gharib Siah Reyhani showed the highest grain elongation (2.04), and Fuji Minori had the widest grains (3.02mm) but the lowest grain shape (1.74) (Figure 1). These differences underscore the diversity in grain quality traits among the studied genotypes. In addition, results identified genotypes matching Iranian consumer preferences. For example, Deylamani (amylose content = 20.90%, gelatinization temperature = 3.75, grain length = 7.33mm, grain shape = 3.54), Hashemi (20.50%, 3.33, 7.03mm, 3.70), and Salari (20.70%, 4.00, 6.88mm, 3.62) exhibited moderate amylose content (19–23%) and gelatinization temperature (3–4), with grain lengths close to or above 7mm and slender grain shapes (≥ 3), although direct consumer preference was not assessed in this study.

Table 2. Descriptive statistics of investigated traits in 48 studied rice genotypes. SE, standard error

Traits	Range	Minimum	Maximum	Mean±SE
Amylose content (%)	12.71	15.97 (Dcl)	28.68 (Saleh)	22.876 ± 0.207
Gelatinization temperature	3.87	3.13 (Hashemi)	7 (Saleh)	4.949 ± 0.086
Grain length (mm)	3.32	5.01 (Fuji Minori)	8.33 (Gohar)	7.173 ± 0.054
Grain width (mm)	1.52	1.54 (Saleh)	3.06 (Fuji Minori)	2.163 ± 0.023
Cooked grain length (mm)	5.9	8.47 (Champa Boodar)	14.37 (Gohar)	11.533 ± 0.079
Grain shape	3.78	1.7 (Fuji Minori)	5.48 (Saleh)	3.428 ± 0.052
Grain elongation	0.73	1.36 (Dorfak)	2.09 (Gharib Siah Reyhani)	1.616 ± 0.010

Table 3. Analysis of variance for grain quality traits among the 48 rice genotypes. * and ** are significant at 0.05 and 0.01 probability, respectively.

Source of Variation	Degrees of Freedom	Mean squares of traits						
		Amylose content (%)	Gelatinization temperature	Grain length (mm)	Grain width (mm)	Cooked grain length (mm)	Grain shape	Grain elongation
Year	1	0.04	0.000	0.008	0.0115	0.001	0.00	0.0003
Varieties	47	63.05**	11.205**	4.450**	0.8500**	8.752**	4.261**	0.14538**
Replication	3	2.15	0.039	0.202	0.0273	0.174	0.031	0.00041
Residuals	332	0.56	0.042	0.091	0.0074	0.249	0.033	0.00677

Table 4. Mean performance for quality-related traits of the 48 studied genotypes with Tukey's honest significant difference (HSD) groupings. Means followed by the same letter within a column are not significantly different at $p < 0.05$ according to Tukey's HSD test.

Genotype	Amylose content (%)	Gelatinization temperature	Grain length (mm)	Grain width (mm)	Cooked grain length (mm)	Grain shape	Grain elongation
Abji Boji	20.50 fg	3.92 ef	7.10 cde	2.05 def	11.50 def	3.46 cd	1.62 bc
AhlamiTarom	21.70 ef	3.33 f	7.57 bcd	1.95 efg	11.73 cde	3.88 bc	1.55 cd
Ali Kazemi	17.30 hi	3.75 ef	7.18 cde	2.22 cd	11.92 bcd	3.23 de	1.66 bc
Anam	25.00 bcd	3.75 ef	7.58 bcd	1.98 efg	12.16 bc	3.82 bc	1.60 bc
Anbarboo	22.90 de	4.92 cd	6.52 efg	2.45 bc	11.17 efg	2.66 fg	1.71 b
Anbarboo Ilam	23.60 cd	4.33 de	6.62 ef	2.37 bc	10.97 fg	2.79 f	1.66 bc
Bejar	26.30 abc	6.75 a	6.83 def	2.07 def	10.60 ghi	3.30 de	1.55 cd
Binam	22.10 e	4.25 de	6.73 def	2.35 bc	10.70 gh	2.86 f	1.59 bc
Champa Boodar	26.90 ab	4.80 cd	5.85 h	2.78 a	8.77 j	2.10 h	1.50 d
Cy	16.80 i	3.83 ef	7.90 abc	2.50 b	12.23 bc	3.16 e	1.55 cd
Dcl	16.70 i	3.66 f	7.95 ab	1.85 gh	12.08 bc	4.30 ab	1.52 d
Deylamani	20.90 f	3.75 ef	7.33 cde	2.07 def	11.17 efg	3.54 c	1.52 d
Dom Sefid	22.50 e	5.50 bc	7.62 bcd	1.88 fg	11.58 def	4.05 b	1.52 d
Dom Siah	22.50 e	5.25 bc	7.15 cde	1.93 efg	11.47 def	3.70 bc	1.60 bc
Dom Zard	21.30 f	3.75 ef	6.30 fg	1.97 efg	10.82 gh	3.20 e	1.72 b
Dorfak	24.40 cd	4.17 de	8.05 ab	1.95 efg	11.48 def	4.13 b	1.43 e
Fajr	23.60 cd	7.00 a	7.58 bcd	1.82 gh	11.93 bc	4.16 b	1.57 bc
Fuji Minori	21.90 ef	5.58 b	5.25 i	3.02 a	9.13 ij	1.74 i	1.74 b
Gharib	20.90 f	4.17 de	6.05 gh	2.67 ab	11.57 def	2.27 gh	1.91 a
Gharib Siah Reyhani	19.60 g	4.25 de	6.27 fg	2.92 a	12.80 ab	2.15 h	2.04 a
Ghodsi	18.90 h	4.66 cd	6.60 ef	2.07 def	11.97 bc	3.19 e	1.81 ab
Gohar	24.20 cd	7.00 a	8.33 a	1.92 fg	14.08 a	4.34 ab	1.69 b
Hashemi	20.50 fg	3.33 f	7.03 de	1.90 fg	11.13 efg	3.70 bc	1.58 bc
Hassan Saraei	21.90 ef	5.16 bc	7.35 cde	2.02 ef	12.03 bc	3.64 c	1.64 bc
Hassani	19.50 g	5.42 bc	6.58 ef	2.60 ab	11.57 def	2.53 g	1.79 ab
IR36	26.20 abc	7.00 a	6.95 def	2.10 de	11.40 def	3.31 de	1.64 bc
IR64	22.30 e	6.00 b	5.68 hi	2.72 a	11.08 fg	2.09 h	1.95 a
Kadous	25.00 bcd	4.08 de	7.98 ab	1.85 gh	11.80 bcd	4.31 ab	1.48 de
Keshvari	23.70 cd	4.25 de	7.70 bc	1.73 h	11.38 ef	4.45 a	1.48 de
Khazar	23.40 cd	4.92 cd	7.35 cde	2.08 def	10.50 hi	3.53 c	1.43 e
Koohsar	22.90 de	4.17 de	6.73 def	2.55 b	10.52 hi	2.64 fg	1.56 cd
Line7	27.00 ab	7.00 a	7.75 bc	1.92 fg	13.28 a	4.04 b	1.71 b
Mohammadi Chaparsar	26.90 ab	4.00 e	6.12 gh	2.48 bc	9.35 i	2.47 g	1.53 d
Mutant Gohar	22.00 ef	3.50 f	7.58 bcd	1.97 efg	11.60 def	3.85 bc	1.53 d
Neda	26.00 abc	4.50 d	7.88 abc	2.07 def	12.20 bc	3.81 bc	1.55 cd
Nemat	26.00 abc	7.00 a	8.63 a	1.95 efg	12.60 abc	4.43 a	1.46 e
Rash	22.60 e	3.83 ef	6.73 def	2.13 d	9.82 hi	3.16 e	1.46 e
Salari	20.70 f	4.00 e	6.88 def	1.90 fg	11.73 cde	3.62 c	1.71 b
Saleh	27.10 a	7.00 a	8.32 a	1.57 i	12.87 ab	5.30 a	1.55 cd
Sange Tarom	22.70 de	5.80 b	7.23 cde	2.50 b	11.57 def	3.53 c	1.60 bc
Sepidroud	26.20 abc	7.00 a	6.92 def	1.92 fg	10.73 gh	3.60 c	1.55 cd
Shahpasand	25.30 bcd	4.75 cd	8.23 a	2.13 d	12.00 bc	3.86 bc	1.46 e
Shiroudi	25.60 abc	4.75 cd	7.53 bcd	1.88 fg	12.40 ab	4.01 b	1.65 bc
TaromMahalli	21.20 f	5.00 cd	7.27 cde	2.08 def	12.20 bc	3.50 c	1.68 b
Tetep	23.20 cd	5.08 bc	7.17 cde	2.28 cd	10.80 gh	3.14 e	1.51 d
TH1	18.50 h	4.41 de	7.37 cde	2.18 cd	13.47 a	3.38 de	1.83 ab
Tisa	26.40 abc	6.25 b	7.90 abc	2.40 bc	13.03 ab	3.29 de	1.65 bc
Zenith	26.40 abc	7.00 a	7.07 cde	2.13 d	10.67 ghi	3.32 de	1.51 d
Tukey HSD (0.05)	1.35	0.62	0.68	0.18	0.95	0.31	0.14

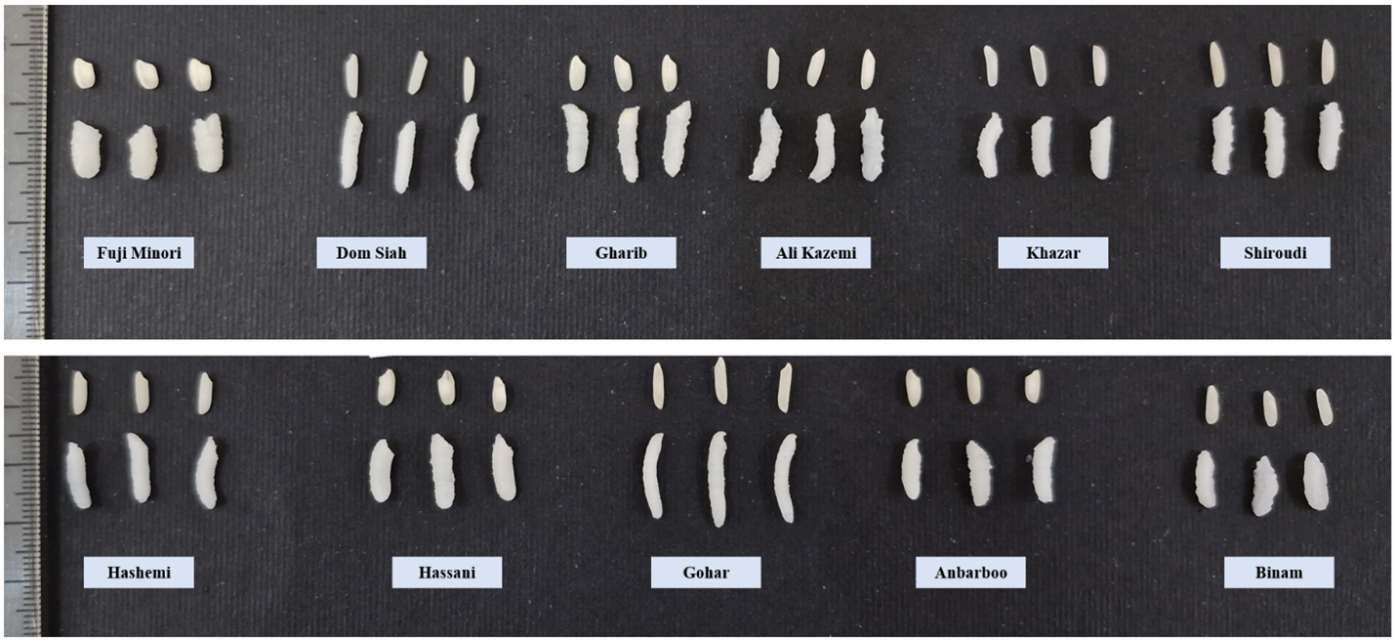


Figure 1. Comparative differences in grain length, grain width, and grain shape, as well as cooked grain length, cooked grain width, and cooked grain shape, among 11 selected rice genotypes (Fuji Minori, Dom Siah, Gharib, Ali Kazemi, Khazar, Shiroudi, Hashemi, Hassani, Gohar, Anbarboo, and Binam). The figure illustrates genotypic variation in both raw and cooked grain dimensions and shape, highlighting differences in grain morphology and cooking-related structural changes among contrasting rice types.

Genetic parameters

The high broad-sense heritability estimates ($H^2 \geq 0.968$, Table 5) reflect substantial genotypic contribution to phenotypic variation under the conditions of this study, facilitated by low genotype × year interaction and minimal residual error. However, it should be noted that such extremely high heritability values likely represent upper-bound estimates obtained under highly controlled experimental conditions. The evaluation was conducted at a single location with uniform agronomic management, using pure-line genotypes and laboratory-based grain quality measurements, all of which tend to reduce environmental variance and inflate heritability estimates. Genetic parameters, including genotypic variance, phenotypic variance, broad-sense heritability, and genetic advance, were estimated using genotype means pooled across years. Year and genotype × year interaction effects were treated as environmental sources of variation and were therefore not included in the calculation of genetic parameters. This approach provides estimates of overall genetic potential that are less influenced by year-specific environmental fluctuations and is widely adopted in multi-year crop evaluation studies (Schmidt *et al*, 2019; Rajaprakasam *et al*, 2025). Estimation of genetic parameters revealed high genetic variability for the studied traits (Table 5). Genotypic variance (σ^2g) was highest for amylose content (7.7834) and lowest for grain elongation (0.0175). Phenotypic variance (σ^2p) followed a similar trend, with values ranging from 0.0181 (grain elongation) to 7.9149 (amylose content). Genotypic coefficients of variation (GCV) were highest for gelatinization temperature (24.018%) and grain shape (21.0828%), indicating substantial genetic diversity, while grain elongation showed the lowest GCV (8.1926%). Phenotypic coefficients of variation (PCV) were slightly higher than GCV, reflecting environmental influences. Broad-sense heritability (H^2) was high for all traits, ranging from 0.968 (grain elongation) to 0.9957 (gelatinization temperature), indicating that trait variation was predominantly genetic. Genetic advance (GA) as a percentage of the mean was highest for gelatinization temperature (49.3704%) and grain shape (43.2934%), suggesting strong potential for selection-based improvement.

Table 5. Estimation of genetic parameters for grain quality related traits of studied rice genotypes. Grand Mean, Arithmetic mean of the population for the respective trait; σ^2g, Genotypic variance; σ^2e, Environmental (error) variance; σ^2p, Phenotypic variance; GCV, Genotypic coefficient of variation (%); PCV, Phenotypic coefficient of variation (%); H^2, Broad-sense heritability; GA, Expected genetic advance under selection; GA Percent Mean, Genetic advance expressed as a percentage of the grand mean.

Trait	Grand Mean	σ^2g	σ^2e	σ^2p	GCV	PCV	H2	GA	GA Percent Mean
Amylose content	22.9104	7.7834	0.1315	7.9149	12.1774	12.2798	0.9834	5.6992	24.8761
Gelatinization temperature	4.9494	1.4131	0.0061	1.4192	24.018	24.0699	0.9957	2.4435	49.3704
Grain length	7.1726	0.5231	0.0125	0.5357	10.084	10.2042	0.9766	1.4724	20.5286
Grain width	2.1635	0.1026	0.0011	0.1036	14.804	14.8806	0.9897	0.6564	30.339
Cooked grain length	11.5327	1.0771	0.0337	1.1108	8.9991	9.1387	0.9697	2.1053	18.255
Grain shape	3.4278	0.5223	0.0033	0.5256	21.0828	21.1497	0.9937	1.484	43.2934
Grain elongation	1.6159	0.0175	6.00E-04	0.0181	8.1926	8.327	0.968	0.2683	16.6044

Correlation among traits

Pearson correlation coefficients revealed significant relationships among traits (Figure 2). Grain length was strongly positively correlated with grain shape ($r = 0.890$, $p < 0.01$) and cooked grain length ($r = 0.675$, $p < 0.01$), but negatively correlated with grain width ($r = -0.711$, $p < 0.01$) and grain elongation ($r = -0.518$, $p < 0.01$). Grain width showed a strong negative correlation with grain shape ($r = -0.918$, $p < 0.01$) and a positive correlation with grain elongation ($r = 0.491$, $p < 0.01$). Amylose content was positively correlated with gelatinization temperature ($r = 0.557$, $p < 0.01$), indicating a strong positive correlation, and showed a weak positive correlation with grain shape ($r = 0.222$, $p < 0.05$), but was negatively correlated with grain elongation ($r = -0.385$, $p < 0.01$). Cooked grain length was positively correlated with grain shape ($r = 0.556$, $p < 0.01$) and grain elongation ($r = 0.276$, $p < 0.05$). These correlations suggest complex interdependencies among traits, influencing grain quality outcomes.

Hierarchical clustering and heatmap

Hierarchical clustering, visualized through a heatmap, grouped the 48 genotypes based on their trait profiles (Figure 3). Four major genotype clusters emerged. Cluster 1 included Gharib Siah Reyhani, IR64, Gharib, and Hassani. This cluster was characterized by high grain elongation and high grain width, with Gharib Siah Reyhani showing the highest grain elongation (2.04). Cluster 2 comprised Fuji Minori, Champa Boodar, and Mohammadi Chaparsar, which also exhibited elevated grain elongation and width, though to a slightly lower extent than Cluster 1. Cluster 3 consisted of Gohar, Line7, Nemat, and Saleh. Within this cluster, Saleh and Line7 displayed high amylose content (27.1% and 27.0%, respectively) and pronounced grain shape (5.3 and 4.04). Cluster 4 was the largest group, containing the remaining genotypes from Dcl to Tarom Mahalli (including Ahlami Tarom, Deylamani, Hashemi, Shiroudi, Dom Sefid, Sange Tarom, and many others). This cluster was associated with intermediate trait values, particularly for cooked grain length and grain length. Trait clustering showed that grain length, grain shape, and cooked grain length were closely related, while grain elongation and grain width formed a separate cluster, consistent with correlation results.

Principal component and genotype biplot analysis

PCA biplot (Figure 4) revealed that PC1 and PC2 explained 49.08% and 24.62% of the total variation, respectively (total 73.70% of the variation). Grain length, grain shape, and cooked grain length were positively associated along PC1, while grain width and elongation negatively correlated with

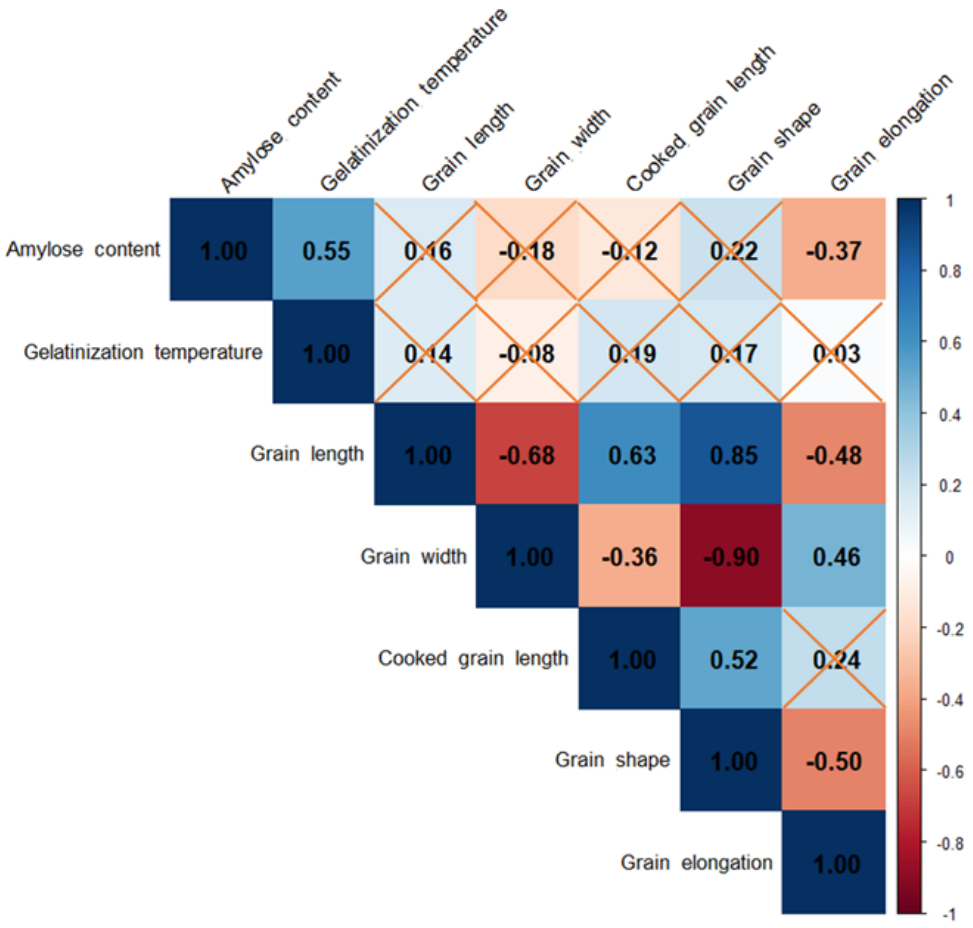


Figure 2. Pearson correlation coefficients among traits in the examined genotypes are represented. Positive correlations are shown in blue, while negative correlations are depicted in red. The multiplication sign (×) denotes non-significant results at the 5% probability level.

these traits. Amylose content and gelatinization temperature were closely aligned along PC2, reflecting their positive correlation (r = 0.557). Variables such as cooked grain length and grain shape exhibited the highest contributions, as indicated by their red-coloured, longer arrows, strongly influencing Dim1 and suggesting a positive correlation with this component. In contrast, variables like grain width and grain elongation showed lower contributions, represented by shorter blue arrows, indicating a weaker influence on the principal components. These findings highlight the key morphometric and physicochemical traits driving variation among the rice genotypes, providing insights into their interrelationships and potential implications for breeding programmes aimed at improving grain quality traits.

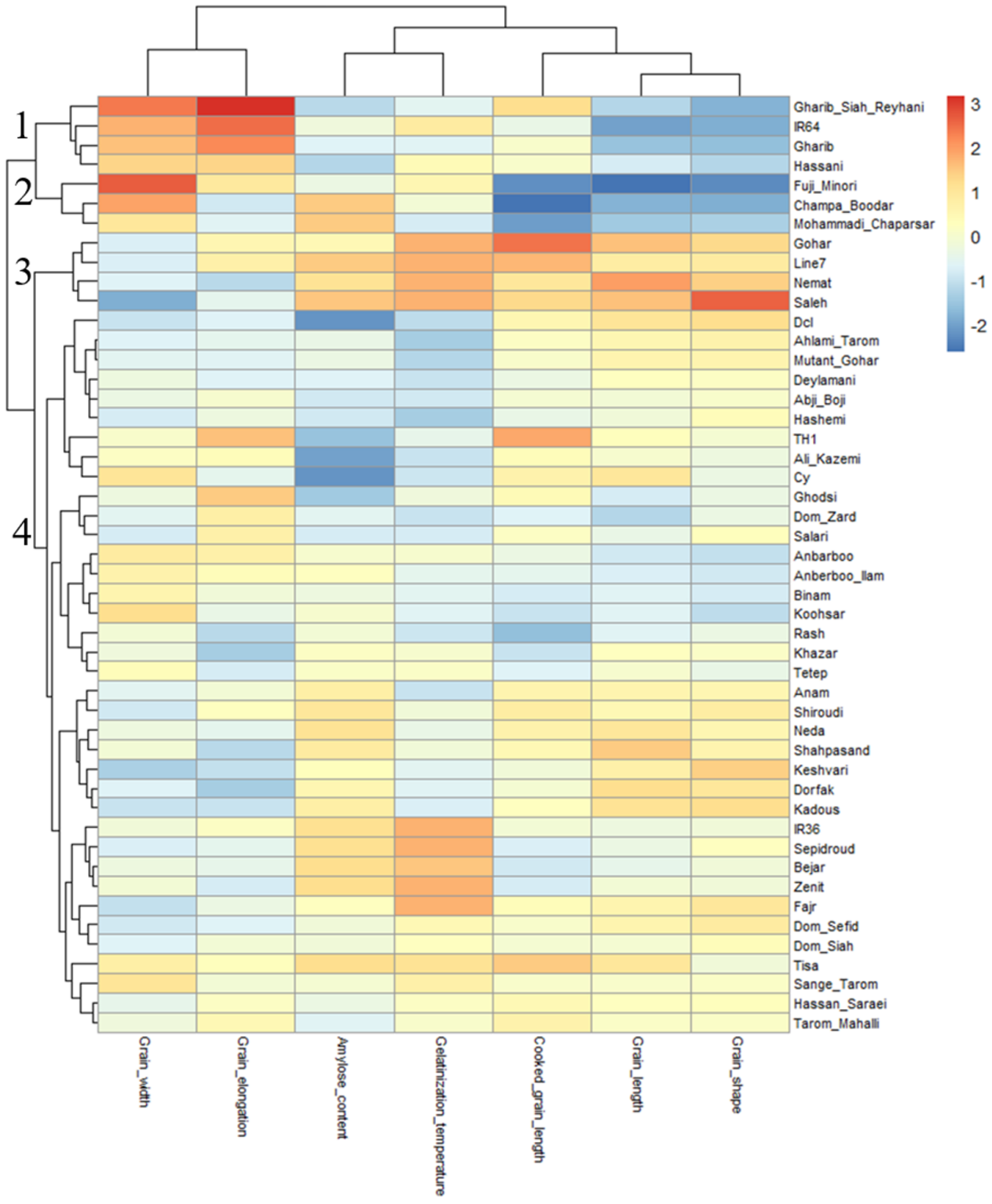


Figure 3. Heatmap with hierarchical clustering of 48 rice genotypes based on seven grain quality traits. Colour intensity represents standardized trait values (blue: low, red: high).

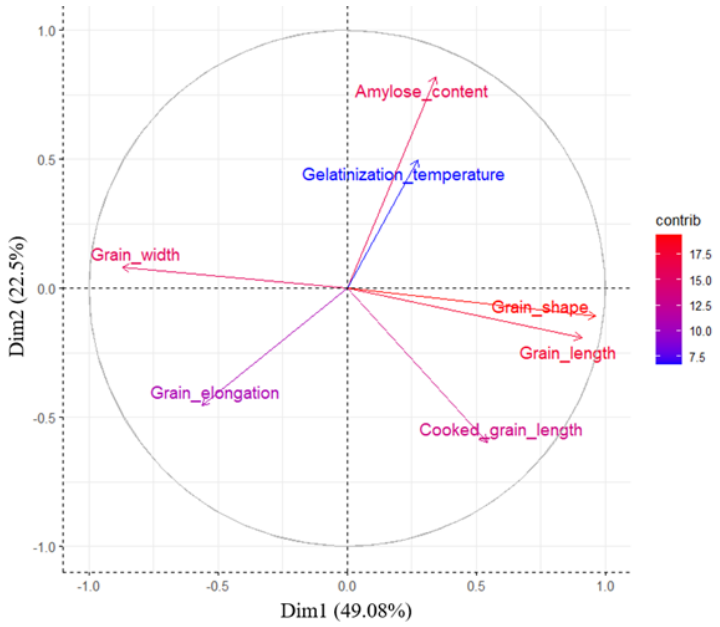


Figure 4. PCA biplot of PC1 vs. PC2 showing the distribution of trait vectors for 48 rice genotypes. PC1 and PC2 explain 49.08% and 24.62% of the total variation, respectively.

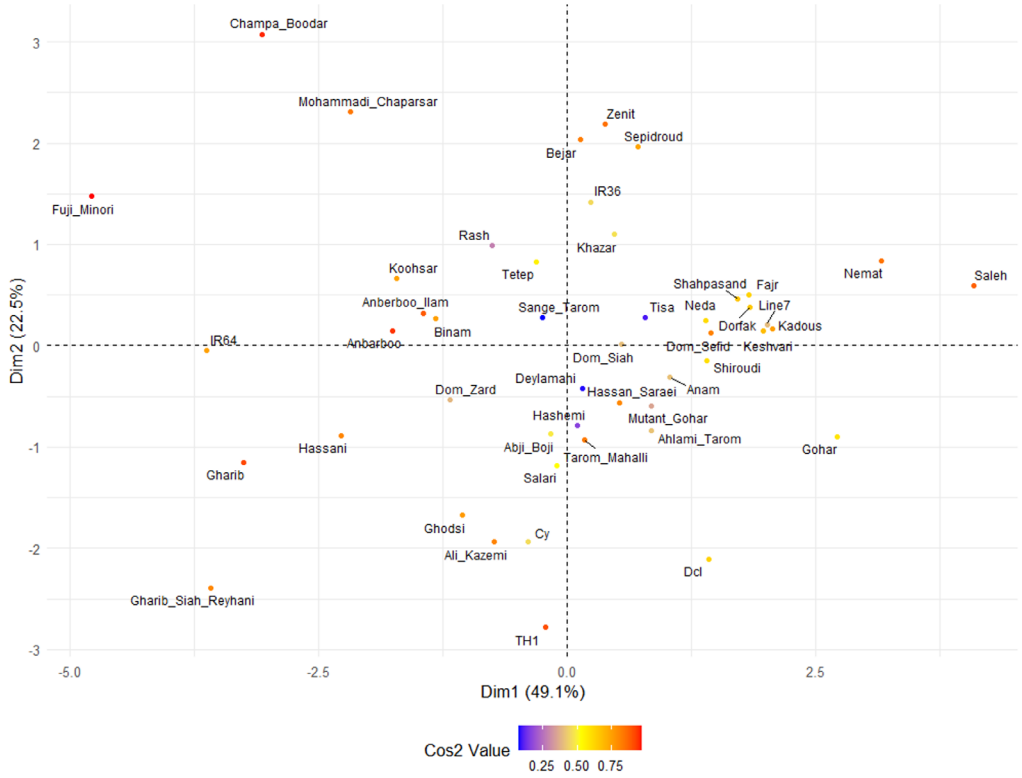


Figure 5. Genotype biplot (cos² colour-coded) of 48 rice genotypes, illustrating genotypic relationships based on grain quality traits. Colour intensity indicates cos² values (purple (low cos², ~0.0-0.25) for minimal contribution, yellow (moderate cos², ~0.25-0.50) for intermediate influence, orange (high cos², ~0.50-0.75) for significant impact, and red (very high cos², ~0.75-1.0) for the strongest association with the analyzed traits).

The genotype biplot (Figure 5), based on cos² values, further elucidated genotypic relationships. Genotypes like Saleh, Nemat, and Saleh were positioned in the upper-right quadrant, associated with high grain shape and cooked grain length. Gharib Siah Reyhani, Hassani, and TH1 clustered in the lower-left quadrant, linked to high grain elongation. Champa Boodar and Fuji Minori were located in the upper-left quadrant of the PCA biplot, reflecting wider grains and more compact shapes compared to the typical long-grain Iranian types. In contrast, genotypes such as Deylamani, Hashemi, and Shiroudi were positioned near the centre of the plot, indicating more balanced trait profiles in terms of grain shape and width.

Path analysis

Before presenting numerical results, it is important to emphasize that the specified path models represent biologically informed hypotheses about trait dependencies rather than definitive causal proofs. The models were designed to reflect the temporal and mechanistic hierarchy of rice grain traits, progressing from primary morphological attributes (grain length and width), to composite descriptors (grain shape), to cooking-induced outcomes (cooked grain length and grain elongation), and finally to key physicochemical determinants of eating quality (amylose content). For a comprehensive understanding of the interrelationships among rice quality characteristics, the selection of appropriate dependent (or resultant) traits for path analysis is paramount. Based on their significant contribution to overall rice quality and consumer preference, cooked grain length and amylose content were chosen as separate dependent variables for independent path analyses. Cooked grain length was selected as a key dependent trait due to its direct relevance to the final physical quality of the grain after cooking. As a primary component of cooking and eating quality, particularly in markets that highly value long, slender grains, cooked grain length serves as an integrated outcome variable. It inherently reflects the combined effects of the raw grain's morphological dimensions and its complex cooking behaviour. Analyzing the direct and indirect influences on cooked grain length allows for a mechanistic understanding of how initial grain properties translate into the ultimate consumer experience.

Concurrently, amylose content was chosen as a second, distinct dependent variable. Despite being a chemical component, amylose content is widely recognized as the single most important determinant of rice cooking and eating quality, profoundly influencing texture attributes such as stickiness, hardness and fluffiness. By modelling amylose content as a dependent trait, the path analysis can elucidate the genetic and morphological factors that underpin this fundamental chemical property.

Cooked grain length: The path analysis (Figure 6A) revealed significant relationships among grain characteristics and cooking properties with cooked grain length as a focal trait. Significant paths were identified. Cooked grain length was strongly influenced by grain length (path coefficient $\beta = 0.699$, $p < 0.001$) but not grain width ($\beta = 0.109$, $p = 0.152$), with an R^2 of 0.398, indicating moderate predictability. Grain shape was highly predicted by grain length ($\beta = 0.454$, $p < 0.001$) and grain width ($\beta = -0.588$, $p < 0.001$), yielding an R^2 of 0.913. Grain elongation was significantly affected by cooked grain length ($\beta = 0.685$, $p < 0.001$) and grain shape ($\beta = -0.853$, $p < 0.001$), with an R^2 of 0.587. Amylose content negatively impacted grain width ($\beta = -0.192$, $p = 0.023$) but not grain length ($\beta = 0.114$, $p = 0.180$). Gelatinization temperature showed no significant effects on grain length ($\beta = 0.078$, $p = 0.357$) or grain width ($\beta = 0.020$, $p = 0.809$). A significant covariance was observed between amylose content and gelatinization temperature ($\beta = 0.545$, $p < 0.001$). These findings highlight cooked grain length as a key mediator between grain length and elongation, with strong interdependencies among grain shape and elongation. This confirms that elongation is primarily a function of post-cooking kernel expansion rather than an independent morphological trait.

Amylose content: Path analysis (Figure 6B) was conducted to elucidate the relationships among grain traits, with amylose content as the central trait. Standardized path coefficients revealed significant ($p < 0.05$) direct effects, with gelatinization temperature positively influencing amylose content ($\beta = 0.52$, $p = 0.002$), indicating that higher gelatinization temperatures enhance amylose content. Grain shape also had a significant positive effect on amylose content ($\beta = 0.127$, $p = 0.008$), suggesting that slender grains correlate with higher amylose levels. Grain shape was strongly determined by grain length ($\beta = 0.454$, $p < 0.001$) and grain width ($\beta = -0.588$, $p = 0.015$), with longer, narrower grains resulting in higher shape ratios. Amylose content significantly influenced cooked grain length ($\beta = -0.228$, $p = 0.012$) and grain elongation ($\beta = -0.446$, $p = 0.020$), highlighting its role in post-cooking grain characteristics. Grain length indirectly affected cooked grain length, while cooked grain length influenced grain elongation ($\beta = 0.845$, $p = 0.005$). Significant covariances were observed between gelatinization temperature and grain shape ($r = 0.058$, $p = 0.030$), and grain length and grain width ($r = -0.676$, $p = 0.008$), indicating interrelated trait dynamics. R-squared values indicated that the model explained 68% of the variance in amylose content. These results underscore amylose content as a pivotal mediator linking physicochemical and morphological traits to cooking quality in rice.

Discussion

The present study provides a comprehensive analysis of grain quality traits in 48 diverse rice genotypes, revealing significant variability, trait interdependencies and genotypic diversity that have important implications for rice breeding programmes aimed at improving cooking and eating quality. The results demonstrate a strong genetic control of grain quality traits across the two growing seasons, as the non-significant year effect from the ANOVA analysis suggests consistent trait expression despite annual variations, while the highly significant genotypic effects emphasized the predominant role of genetic diversity among the evaluated rice genotypes. The negligible replication effects further indicate minimal environmental interference, reinforcing that observed differences are genetically driven, which is advantageous for targeted breeding programmes aimed at enhancing desirable traits.

The stability of rice grain quality traits across different years observed in our study is supported by Kunjaroenruk *et al* (2025), who reported that certain rice quality parameters, evaluated across multiple seasons in a tropical savannah environment, remained unaffected by seasonal variations, highlighting the predominant role of genotypic factors over temporal environmental changes. The significant genotypic variation observed across all traits (Table 3) aligns with previous studies on rice grain quality, such as those by Ferdous *et al* (2018) and Karim *et al* (2024), which reported substantial diversity in amylose content, grain dimensions and cooking properties among rice cultivars. The broad variation observed among genotypes reflects the inclusion of Iranian landraces, improved varieties, and lines, providing a rich genetic pool for selection. This diversity is critical for breeding programmes targeting specific quality traits, such as high grain elongation for consumer-preferred aromatic varieties or tailored amylose levels for different cooking applications. The HSD test at the

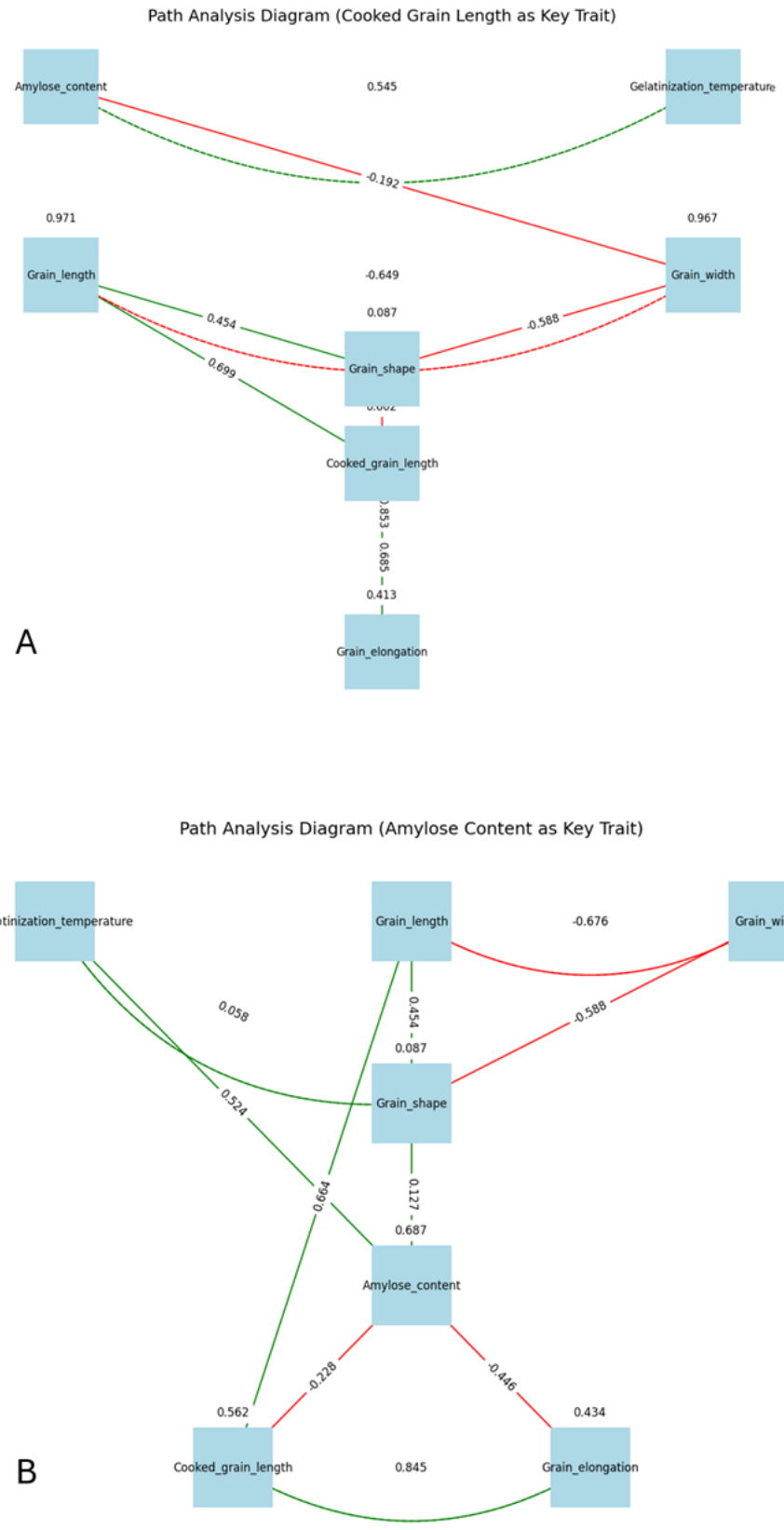


Figure 6. Path analysis diagram. A, Cooked grain length as key trait. B, Amylose content as key trait. The numbers on the connecting lines represent path coefficients (β), which indicate the strength and direction of the relationships between traits: positive values signify a positive influence, while negative values signify a negative influence. The magnitude of these coefficients reflects the degree of impact. The colours of the lines denote the type of relationship: green lines represent positive effects, red lines indicate negative effects, and the thickness of the lines corresponds to the relative strength of the coefficients.

5% level (Table 4) delineated genotypic differences, and high broad-sense heritability and genetic advance highlight traits with strong potential for selection in breeding programmes.

The high broad-sense heritability estimates and genetic advance values indicate that many of the evaluated grain quality traits possess strong potential for improvement through selection. However, these exceptionally high heritability estimates should be interpreted with caution, as grain quality traits are generally more stable and less environmentally sensitive than yield-related traits, particularly when measured under controlled laboratory conditions. Moreover, the absence of significant genotype × year interaction, combined with single-location testing, uniform crop management, and the use of genetically fixed pure lines, likely reduced environmental variance and inflated heritability estimates. Such inflation under optimized experimental conditions has been widely reported in rice quality studies and does not necessarily imply similar expression across diverse environments or production systems. Therefore, the heritability estimates reported here should be regarded as indicators of genetic potential under favourable and controlled conditions rather than absolute parameters. Similar observations have been reported by Nirmaladevi *et al* (2015), Hori and Sun (2022) and Al-Daej (2022), who emphasized the relative stability of grain quality traits compared with yield-related traits.

Correlation analysis (Figure 2) revealed important relationships among grain quality traits that have direct implications for breeding strategies. The strong positive associations among grain length, grain shape and cooked grain length highlighted grain length as a key morphometric trait influencing visual appeal and cooking quality, as supported by previous literature (Anne *et al*, 2018; Arikit *et al*, 2019; Cruz *et al*, 2021). Conversely, the negative associations between grain length and grain elongation, as well as between grain shape and grain elongation, suggest potential trade-offs, where slender or longer grains may exhibit reduced elongation during cooking. This phenomenon may be attributed to structural constraints in the starch matrix and has been reported previously for premium rice types such as Basmati (Bhattacharjee *et al*, 2002; Singh *et al*, 2018). The substantial inclusion of Iranian landraces in this study revealed unique grain quality profiles – such as the outstanding elongation of Gharib Siah Reyhani and the preferred moderate amylose and gelatinization temperature combinations in Deylamani, Hashemi, and Salari – that are characteristic of traditional Iranian rice types valued by consumers. These findings emphasize the uniqueness of landrace germplasm and its irreplaceable role as a representative reservoir of genetic variation for grain quality.

The positive association between amylose content and gelatinization temperature is consistent with the biochemical properties of starch, as higher amylose levels generally require greater thermal energy for gelatinization. Similar relationships have been documented by Xu *et al* (2024) and Zhao *et al* (2022b). The observed negative association between amylose content and grain elongation suggests that high-amylose genotypes may be less suitable for rice types where pronounced elongation after cooking is preferred, such as aromatic rice. Comparable findings were reported by Karim *et al* (2024), although contrasting results have been observed by Bandara *et al* (2025), indicating that genotype composition and genetic background strongly influence these relationships. The substantial inclusion of Iranian landraces in this study revealed distinctive grain quality profiles – such as superior elongation in certain traditional genotypes and favourable combinations of moderate amylose content and gelatinization temperature in others – that are characteristic of rice types preferred by Iranian consumers. These results highlight the importance of landrace germplasm as an irreplaceable reservoir of genetic diversity for grain quality improvement, particularly for region-specific breeding objectives.

Hierarchical clustering (Figure 3) effectively grouped genotypes based on their overall grain quality profiles, allowing the identification of clusters with distinct and agronomically relevant characteristics. Genotypes characterized by high grain elongation and width may serve as promising parents for breeding programmes targeting elongation, while those with high amylose content and slender grain shape may be suitable for markets requiring firmer, non-sticky cooked rice. The largest cluster, comprising genotypes with intermediate trait values, reflects broad adaptability and potential versatility for diverse consumer preferences. The clustering of traits further reinforced the strong interrelationships among grain length, grain shape and cooked grain length, consistent with correlation and path analysis results. Similar trait-based clustering patterns have been reported by Abdelsalam *et al* (2025).

Principal component analysis serves as a powerful tool for identifying trends, reducing redundancy and elucidating complex interrelationships within datasets, even across diverse crop species characterized by varying yields and grain attributes (Islam *et al*, 2024). This versatility underscores its utility in agricultural research, where multifaceted traits must be analyzed to inform breeding decisions. By focusing on maximizing the explained variance while accounting for trait interconnections, PCA facilitated the classification of accessions into distinct groups, highlighting genotypic diversity and potential trade-offs in quality profiles (Khatun *et al*, 2023). PCA and genotype biplots (Figures 3 and 4) provided deeper insight into trait associations and varietal performance. Along PC1, which accounted for 49.08% of the variance, positive loadings were observed for grain length, grain shape and cooked grain length, indicating their coordinated contribution to overall grain morphology and cooking appeal. In contrast, grain width and grain elongation exhibited negative associations, consistent with correlation analyses that suggest antagonistic relationships – longer grains often correspond to reduced width, potentially impacting milling efficiency and breakage susceptibility. This mirrors observations in Abdelsalam *et al* (2025), where PCA revealed two primary dimensions explaining 94.2% of variance: one dominated by physical-processing traits and another by nutritional-functional attributes. Furthermore, the alignment of amylose content and gelatinization temperature along PC2 (22.5% variance) reinforces their biochemical interdependence, as amylose influences starch granule stability and gelatinization behaviour during cooking. This association aligns with Cruz *et al* (2021), who reported strong genetic correlations among amylose content, gelatinization temperature (measured as alkali-spread value), and pasting properties in Latin American rice germplasm.

The path analysis underscored cooked grain length's pivotal role in rice quality, with grain length strongly predicting it, aligning with Devi *et al* (2017), and John and Raman (2023)

who found uncooked grain length as a primary determinant of cooked grain length across rice varieties. However, the negative effect of grain shape on elongation contrasts with Devi *et al* (2017), who reported a positive relationship, possibly due to genotypic diversity between their samples and ours. These results indicate that amylose content has a significant but limited effect on grain width, while gelatinization temperature has no significant effect on grain dimensions. In contrast, Bhardwaj *et al* (2019) suggested that higher amylose content is associated with longer and broader grains in indica rice.

Amylose content is a critical mediator in rice grain quality, significantly influenced by gelatinization temperature and grain shape, and affecting cooked grain length and grain elongation. The positive effect of gelatinization temperature on amylose content aligns with findings by Jane *et al* (2025), who reported that gelatinization temperature directly influences amylose content by promoting amylose leaching and inducing structural changes in starch granules. In addition, Neoh *et al* (2020) and Yang *et al* (2024) reported a positive correlation between higher gelatinization temperatures and higher amylose content. However, this result contrasts with Pang *et al* (2016), who reported that correlation analyses revealed a high amylose content in inbred lines was associated with lower gelatinization temperatures, suggesting varietal differences may account for the discrepancy. The significant influence of grain shape on amylose content supports Al-Daej (2022), who found significant positive correlations between grain shape and amylose content.

Conclusion

In this study, we characterized 48 rice genotypes for key grain quality traits, revealing substantial genetic variation and high broad-sense heritability for traits such as grain shape, gelatinization temperature and grain length, indicating strong potential for selection in quality-oriented breeding programmes. Our findings identified genotypes with favourable trait combinations, including long, slender grains with moderate amylose content, which, according to prior literature, align with preferred rice characteristics in Iranian and regional markets. However, this study has several limitations: (1) evaluations were conducted at a single location, which may limit the generalizability of the results across diverse environments; (2) the data are purely phenotypic, without molecular or genomic validation; (3) genetic analyses were restricted to broad-sense heritability, which does not partition additive and non-additive effects; and (4) multi-environment performance and genotype × environment interactions were not assessed. Despite these limitations, the study provides a robust framework for initial phenotypic selection of high-quality rice genotypes and highlights promising candidates for future breeding. Further research incorporating multi-location trials, molecular marker data and advanced genetic analyses is recommended to validate and extend these findings.

Supplemental data

Supplemental File 1. Python Script for Path Analysis Diagram of Cooked Grain Length as a Key Trait.

Data availability

The datasets generated during the current study are available from the corresponding author on reasonable request.

Acknowledgments

This work was supported by the Rice Research Institute of Iran. The authors thank Mr Mohammad Mohammadi, Mrs Mitra Yekta, Dr Mina Ebrahimi, Mrs Tahereh Mohseni, and Mr Mahmoud Sayyadi for their assistance and companionship.

Conflict of interest statement

The author declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as potential conflicts of interest.

References

Abarshahr, M., Rabiei, B., Lahigi, H.S. (2011). Genetic variability, correlation and path analysis in rice under optimum and stress irrigation regimes. *Notulae Scientia Biologicae* 3 (4), 134-142. https://doi.org/10.15835/nsb346280

Abdelsalam, K.M., Shaalan, A.M., AbouEl-Soud, G.M., El-Dalil, M.A., Marei, A.M., El-Moneim, D.A., El-Banna, A.A., Lamlom, S.F., Abdelghany, A.M. (2025). Comprehensive quality profiling and multivariate analysis of rice (*Oryza sativa* L.) cultivars: integrating physical, cooking, nutritional, and micronutrient characteristics for enhanced varietal selection. *BMC Plant Biology* 25 (1), 492. https://doi.org/10.1186/s12870-025-06438-5

Al-Daej, M.I. (2022). Genetic studies for grain quality traits and correlation analysis of mineral element contents on Al-Ahsa rice and some different varieties (*Oryza sativa* L.). *Saudi Journal of Biological Sciences* 29 (3), 1893-1899. https://doi.org/10.1016/j.sjbs.2021.10.032

Anne, S., Islam, S., Haque, M., Raffi, S. (2018). Evaluation of Physical and Cooking Property Based Grain Quality Traits in Popular Rice (*Oryza sativa* L.) Cultivars of Bangladesh. *Sch. J. Agric. Vet. Sci* 5 (3), 172-178.

Arikit, S., Wanchana, S., Khanthong, S., Saensuk, C., Thianthavon, T., Vanavichit, A., Toojinda, T. (2019). QTL-seq identifies cooked grain elongation QTLs near soluble starch synthase and starch branching enzymes in rice (*Oryza sativa* L.). *Scientific reports* 9 (1), 8328. https://doi.org/10.1038/s41598-019-44856-2

Bandara, A., Batugahage, B., Prasantha, B., Senanayake, D. (2025). Impact of amylose content on physical properties, cooking qualities, and water absorption of long-grain basmati-type rice (*Oryza sativa* L.) varieties. *Ceylon Journal of Science* 54 (1). https://doi.org/10.4038/cjs.v54i1.8418

Bhardwaj, R., Salgotra, R., Sharma, M. (2019). Studies on correlation of amylose content and grain dimensions in Basmati rice (*Oryza sativa* L.). *Electronic Journal of Plant Breeding* 10 (2), 364-369. http://dx.doi.org/10.5958/0975-928X.2019.00046.2

Bhattacharjee, P., Singhal, R.S., Kulkarni, P.R. (2002). Basmati rice: a review. *International Journal of Food Science and Technology 37* (1), 1-12. https://doi.org/10.1046/j.1365-

2621.2002.00541.x
Birla, D.S., Malik, K., Sainger, M., Chaudhary, D., Jaiwal, R., Jaiwal, P.K. (2017). Progress and challenges in improving the nutritional quality of rice (*Oryza sativa* L.). *Critical Reviews in Food Science and Nutrition* 57 (11), 2455-2481. https://doi.org/10.1080/10408398.2015.1084992
Cruz, M., Arbelaez, J.D., Loaiza, K., Cuasquer, J., Rosas, J., Graterol, E. (2021). Genetic and phenotypic characterization of rice grain quality traits to define research strategies for improving rice milling, appearance, and cooking qualities in Latin America and the Caribbean. *The Plant Genome* 14 (3), e20134. https://doi.org/10.1002/tpg2.20134
Devi, K.R., Chandra, B.S., Lingaiah, N., Hari, Y., Venkanna, V. (2017). Analysis of variability, correlation and path coefficient studies for yield and quality traits in rice (*Oryza sativa* L.). *Agricultural Science Digest-A Research Journal* 37 (1), 1-9. https://doi.org/10.18805/asd.v0iOF.7328
Ferdous, N., Elias, S.M., Howlader, Z.H., Biswas, S.K., Rahman, M.S., Habiba, K.K., Seraj, Z.I. (2018). Profiling Bangladeshi rice diversity based on grain size and amylose content using molecular markers. *Current Plant Biology* 14 56-65. https://doi.org/10.1016/j.cpb.2018.09.002
Hori, K., Sun, J. (2022). Rice grain size and quality. *Rice* 15 (1), 33. https://doi.org/10.1186/s12284-022-00579-z
Islam, M.A., Hasan, M.M., Hossain, M.A., Haque, M.A., Siddique, M.N.A., Shamsuddoha, M., Habib, M.A., Risha, S.S. (2024). Identification and evaluation of high-performing advanced germplasm of rice through morphological and breeding value analysis. *Discover Agriculture* 2 (1), 1-13. https://doi.org/10.1007/s44279-024-00143-x
Jane, K.A., Inamdar, N.N., Kotagale, N.R. (2025). Effect of gelatinizing temperature and moisture on the retrogradation of coix starch. *International Journal of Biological Macromolecules* 143860. https://doi.org/10.1016/j.ijbiomac.2025.143860
John, D., Raman, M. (2023). Physicochemical properties, eating and cooking quality and genetic variability: A comparative analysis in selected rice varieties of South India. *Food Production, Processing and Nutrition* 5 (1), 49. https://doi.org/10.1186/s43014-023-00164-x
Juliano, B., Perez, C. (1984). Results of a collaborative test on the measurement of grain elongation of milled rice during cooking. *Journal of Cereal Science* 2 (4), 281-292. https://doi.org/10.1016/S0733-5210(84)80016-8
Juliano, B.O. (1971). A simplified assay for milled-rice amylose. Cereal Sci. Today 12 334-360. url: https://www.researchgate.net/publication/275886661_A_Simplified_Assay_for_Milled-Rice_Amylose
Karim, M.D., Abuhena, M., Hossain, M.D., Billah, M.M. (2024). Assessment and comparison of cooking qualities and physio-chemical properties of seven rice varieties in terms of amylose content. Food Physics 1 100014. https://doi.org/10.1016/j.foodp.2024.100014
Khatun, M., Islam, M., Khan, M., Hossain, M., Islam, A. (2023). Principal component analysis and estimated breeding values for selecting suitable parental genotypes in rice (*Oryza sativa* L.). *Sabrao Journal of Breeding & Genetics* 55 (6). http://doi.org/10.54910/sabrao2023.55.6.3
Kumar, A.P., Sarawgi, A., Verulkar, S., Verma, R. (2010). Correlation coefficient and path analysis study among grain quality components in rice (*Oryza sativa* L.). *Electronic Journal of Plant Breeding* 1 (6), 1468-1473. https://www.cabidigitallibrary.org/doi/pdf/10.5555/20113103136
Kunjaroenruk, J., Koonmanee, S., Singkham, N., Chankaew, S., Suriharn, K. (2025). Genotypic variation and seasonal effects on rice (*Oryza sativa* L.) grain protein content and yield in tropical Savannah environment. *Journal of Agriculture and Food Research* 20 101778. https://doi.org/10.1016/j.jafr.2025.101778
Little, R.R. (1958). Differential effect of dilute alkali on 25 varieties of milled white rice. *Cereal Chem*. 35 111-126.
Malik, A., Kumar, A., Ellur, R.K., Krishnan S, G., Dixit, D., Bollinedi, H., Vinod, K., Nagarajan, M., Bhowmick, P., Singh, N. (2022). Molecular mapping of QTLs for grain dimension traits in Basmati rice. *Frontiers in Genetics* 13 932166. https://doi.org/10.3389/fgene.2022.932166
Modarresi, M. (2023). Rice Breeding in Iran, Current Status and Future Perspective. *Plant Breeding and Biotechnology* 11 (2), 97-104. https://doi.org/10.9787/PBB.2023.11.2.97
Modarresi, M., AllahGholipour, M., Ebadi, A. (2024). Estimation of Gene Effect and Combining Ability for Yield and Yield Components Using Line x Tester Analysis in Rice (*Oryza sativa* L.). *Plant Breeding and Biotechnology* 12 17-29. https://doi.org/10.9787/PBB.2024.12.17
Neoh, G.S., Tan, X., Dieters, M., Fox, G., Gilbert, R. (2020). Effects of cold temperature on starch molecular structure and gelatinization of late-maturity alpha-amylase affected wheat. *Journal of Cereal Science* 92 102925. https://doi.org/10.1016/j.jcs.2020.102925
Nirmaladevi, G., Padmavathi, G., Suneetha Kota, S.K., Babu, V.R. (2015). Genetic variability, heritability and correlation coefficients of grain quality characters in rice (*Oryza sativa* L.). *SABRAO Journal of Breeding and Genetics* 47 (4), 424-433.
Pang, Y., Ali, J., Wang, X., Franje, N.J., Revilleza, J.E., Xu, J., Li, Z. (2016). Relationship of rice grain amylose, gelatinization temperature and pasting properties for breeding better eating and cooking quality of rice varieties. *PloS one* 11 (12), e0168483. https://doi.org/10.1371/journal.pone.0168483
R-Core-Team (2023). R: a language and environment for statistical computing; R Core team: Vienna, Austria, 2023. Available at: www. r-project. org (Accessed February 17, 2023).
Rajaprakasam, S., Jafarullakhan, S.S., Vijayakumar, V., Saravanan, N.A., Rathinavelu, S., Kannan, B., Chockalingam, V., Muthurajan, R., Kanagarajan, S. (2025). Multi-model-based validation of multi-environment trial results in horse gram (*Macrotyloma uniflorum* Lam. Verdc.). *Scientific Reports* 15 (1), 41266. https://doi.org/10.3390/metabo13020165
Safiedin Ardebili, S.M., Azizinezhad, R., Saadatmand, S., Ahmadvand, R., Mehregan, I. (2024). Investigation of the Morphometric Diversity of Ali Kazemi Rice in Iran. *Journal of Genetic Resources* 10 (2), 131-142. url: https://sc.journals.umz.ac.ir/article_5069_938101953784273f928db45fff441fa5.pdf
Schmidt, P., Hartung, J., Bennewitz, J., Piepho, H.P. (2019). Heritability in plant breeding on a genotype-difference basis. *Genetics* 212 (4), 991-1008. https://doi.org/10.1534/genetics.119.302134
Singh, V., Singh, A.K., Mohapatra, T., S, G.K., Ellur, R.K. (2018). Pusa Basmati 1121–a rice variety with exceptional kernel elongation and volume expansion after cooking. *Rice* 11 1-10. https://doi.org/10.1186/s12284-018-0213-6
Sultana, S., Faruque, M., Islam, M.R. (2022). Rice grain quality parameters and determination tools: a review on the current developments and future prospects. *International Journal of Food Properties* 25 (1), 1063-1078. https://doi.org/10.1080/10942912.2022.2071295
Talekar, S., Praveena, M.V., Satish, R. (2022). Genetic diversity

using principal component analysis and hierarchical cluster analysis in rice. *Int J Plant Sci* 17 (2), 191-196. https://doi.org/10.15740/HAS/IJPS/17.2/191-196

Thuy, N.P., Trai, N.N., Khoa, B.D., Thao, N.H.X., Phong, V.T., Thi, Q.V.C. (2023). Correlation and path analysis of association among yield, micronutrients, and protein content in rice accessions grown under aerobic condition from Karnataka, India. *Plant breeding and biotechnology* 11 (2), 117-129. https://doi.org/10.9787/PBB.2023.11.2.117

Xu, H., Xu, S., Xu, Y., Jiang, Y., Li, T., Zhang, X., Yang, J., Wang, L. (2024). Relationship between the physicochemical properties and amylose content of rice starch in rice varieties with the same genetic background. *Journal of Cereal Science* 118 103932. https://doi.org/10.1016/j.lwt.2026.119398

Yadi, R., Heravan, I.M., Heidari Sharifabad, H. (2021). Identifying the superior traits for selecting the ideotype of rice cultivars. *Cereal Research Communications* 49 475-484. https://doi.org/10.1007/s42976-020-00088-z

Yang, J., Zhang, X., Wang, D., Wu, J., Xu, H., Xiao, Y., Xie, H., Shi, W. (2024). The deterioration of starch physiochemical and minerals in high-quality indica rice under low-temperature stress during grain filling. *Frontiers in Plant Science* 14 1295003. https://doi.org/10.3389/fpls.2023.1295003

Zhao, D., Zhang, C., Li, Q., Liu, Q. (2022a). Genetic control of grain appearance quality in rice. *Biotechnology Advances* 60 108014. https://doi.org/10.1016/j.biotechadv.2022.108014

Zhao, T., Zhang, H., Chen, F., Tong, P., Cao, W., Jiang, Y. (2022b). Study on structural changes of starches with different amylose content during gelatinization process. *Starch-Stärke* 74 (7-8), 2100269. https://doi.org/10.1002/star.202100269

Genetic Resources

REVIEW

Genetic Resources (2026), 7 (13), 169–180
DOI: 10.46265/genresj.OVEO3504
https://www.genresj.org
ISSN: 2708-3764

A comprehensive review of approaches for genetic improvement in foxtail millet (*Setaria italica* L.)

Shivika Pareek, Reginah Pheirim*, C.P. Chetariya, Alka Soharu

Department of Genetics and Plant Breeding, School of Agriculture, Lovely Professional University, Jalandhar, Punjab-144411, India

Abstract: Foxtail millet (*Setaria italica* (L.) P. Beauv.) is a historically significant and resilient cereal crop known for its adaptability to diverse environmental conditions, nutritional benefits and economic potential. Despite its importance, foxtail millet remains underutilized compared to major cereals. Recent advances in breeding techniques and molecular marker technologies have substantially enhanced efforts toward its genetic improvement. Conventional breeding approaches, including selection, hybridization and mutation breeding, have contributed to trait improvement, while modern strategies such as marker-assisted selection (MAS), genome-wide association studies (GWAS), and quantitative trait loci (QTL) mapping have accelerated the identification of genes associated with desirable agronomic traits. The development of high-resolution genetic maps and the use of molecular markers, including simple sequence repeats (SSRs) and single-nucleotide polymorphisms (SNPs), have further facilitated genomic studies and breeding programmes. In addition, association mapping has emerged as an effective approach for identifying genomic regions linked to important agronomic traits, thereby supporting precision breeding. With the integration of genomics, transcriptomics and genome-editing technologies, foxtail millet improvement is progressing towards enhanced yield, stress tolerance and nutritional quality. This review summarizes recent advances in foxtail millet genetics, breeding strategies and molecular marker technologies, highlighting their significance for sustainable agriculture and global food and nutritional security.

Keywords: Foxtail millet, breeding approaches, molecular marker technologies, MAS, GWAS, genome editing

Citation: Pareek, S., Pheirim, R., Chetariya, C. P. and Soharu, A. (2026) "A comprehensive review of approaches for genetic improvement in foxtail millet (*Setaria italica* L.)", *Genetic Resources*, 7(13), pp. 169–180. doi: 10.46265/genresj.OVEO3504.

Introduction

Setaria italica (L.) P. Beauv., commonly known as foxtail millet, is among the earliest domesticated millet species and has a long history of cultivation in India since ancient times (Jia *et al*, 2013). Known locally as 'Kangni' in Hindi, foxtail millet is a hardy crop that is well adapted to a wide range of climatic conditions. It can thrive in both low-fertility and well-drained fertile soils and is commonly cultivated in semi-arid and rainfed regions (Fukunaga and Kawase, 2024). Owing to its ability to grow under low-input conditions, the crop is often regarded as a 'life-saving crop' for farmers in marginal environments (Zhang *et al*, 2022).

Among the minor millets, foxtail millet ranks as the second most important species and is recognized for its high nutritional value. The grain typically contains about 65% carbohydrates, 11% protein and 6% fat, and is rich in minerals such as iron and copper, as well as dietary fibre (Yousaf *et al*, 2021). It serves as a staple food in some regions of Southern India. Due to its immense nutritional value, millet is also considered one of the best weaning foods. In today's modern world, where growing consumer demand for healthier food, foxtail can serve as one of the best choices for the diet/weight-conscious, diabetic and people with heart disease, and holds great potential due to its unique phenolic content and roughage (Tripathi *et al*, 2021). In addition to human consumption, foxtail millet is used in the production of alcoholic beverages and also serves as fodder and feed for livestock in both temperate and tropical regions (Tomar *et al*, 2023).

* Corresponding author: Reginah Pheirim (pheirimreginah4@gmail.com)

Received: 28.08.2025 | Accepted: 24.03.2026 | Published online: 12.05.2026

Despite its resilience, adaptability, and nutritional advantages, foxtail millet remains far less widely cultivated than major cereal crops. Historically, the crop has received limited research attention, leading to its underutilization and its perception as a 'forgotten crop,' often labelled as a 'poor man's food' (Singh *et al*, 2023). In contrast, major cereals such as rice, maize and wheat, which together contribute over 60% of the global plant-based dietary energy supply, are relatively poor sources of essential micronutrients. This heavy reliance on micronutrient-poor staple crops has significantly contributed to the problem of 'hidden hunger' (Vetriventhan *et al*, 2019).

With the rapidly increasing global population and the growing challenges of climate change, there is an urgent need to diversify agricultural systems to ensure sustainable food and nutritional security. In this context, foxtail millet has gained renewed attention due to its climate resilience, nutritional richness and potential role in sustainable agriculture. Therefore, this review provides a comprehensive overview of foxtail millet, with particular emphasis on its genetic resources, breeding strategies and molecular approaches that have been explored for its genetic improvement.

Morphological description of foxtail millet

Foxtail millet is a robust and upright annual grass that typically grows to a height of 0.6–1.2m (Brink, 2006). The plants are mostly slender with stem tillers but may occasionally branch, typically producing 1 to 25 culms, with an average of 3–4. It possesses a well-developed root system, with slender, wiry adventitious roots originating from the lower nodes (Rahayu *et al*, 1996). Leaves are simple and alternate with a glabrous or slightly hairy sheath measuring 10 to 25cm in length. The ligule is short and fimbriate, while the linear leaf blade varies from 15 to 50cm in length and 0.5 to 4cm in width, featuring an acuminate apex and a slightly rough, prominent midrib. It has a short vegetative growth period (Dekker, 2003), ranging from 80 to 120 days, though some cultivars mature within 60 days (Swamy, 2023).

Floral biology

The millet's inflorescence, measuring 6–30cm, consists of a central stalk surrounded by short, bristling, prickly lateral branches (Figure 1B). The terminal spike usually ranges from 8 to 32cm in length with a drooping appearance (Figure 1A), is thick with cylindrical, lobed structure and held up by a very tiny, slender pedicel (Sundararaj and Thulasidas, 1976). It consists of pairs of tiny blooms on each spikelet, which are enclosed by two glumes. The top bloom is fertile and bisexual, while the lower flower is sterile (Nirmalakumari and Vetriventhan, 2010). Each floret consists of three stamens with anthers that are white or yellow in colour (Jayaraman *et al*, 1997). The pistil of the flower consist of a smooth, round ovary with two long styles and feathery stigmas. The fruit is a caryopsis (grain), approximately 2mm in length and oval in shape. The colour varies from pale yellow to orange, red, brown, or black, with the lemma and palea tightly enclosing the grain with a 1,000-seed weight of around 2g (Figure 1C).

Anthesis and pollination

Foxtail millet is a self-pollinating species with a low average natural outcrossing rate of approximately 4%, occasionally spontaneous hybrids forming between wild and cultivated varieties (Till-Bottraud *et al*, 1992). Flowering commences when roughly three-quarters of the head protrudes from the sheath and proceeds to flower in a vertically aligned manner (Sundararaj and Thulasidas, 1976). The flowering process for a head takes between 8 to 16 days. Anthesis commonly

Figure 1. Structure of the foxtail millet plant. (A) Mature plant (B) Inflorescence (C) Grains. Created by the authors.

occurs early in the morning and around midnight, although this can vary widely based on environmental conditions (Siles *et al*, 2001).

Global status of germplasm resources

S. italica is predominantly grown in Asia, parts of Europe, and Africa, playing a vital role in the diets of people in China, India, Korea, Japan and Nepal (Dwivedi *et al*, 2012). Countries such as China, India, France and Japan have the most extensive collections of foxtail millet germplasm (Vetriventhan *et al*, 2016). The Crop Trust, along with other global organizations, is actively engaged in the ex situ and in situ conservation of foxtail millet genetic resources (Bramel *et al*, 2022). The Chinese National Gene Bank (CNGB) preserves a vast collection of 26,670 germplasm accessions (Wang *et al*, 2012). The International Crops Research Institute for the Semi-Arid Tropics (ICRISAT) houses germplasm sourced from 26 different countries, while the Plant Genetic Resources Conservation Unit (PGRCU) in the USA and the National Institute of Agro-biological Sciences (NIAS) in Japan also maintain diverse germplasm collections (Upadhyaya *et al*, 2009; 2011). The formation of representative germplasm, which includes core and mini core collections (Upadhyaya *et al*, 2009; 2011), serves as valuable genetic resources for research at the genomic level.

Cytogenetics studies in foxtail millet

Foxtail millet is a C4 crop with a chromosome number of 2n = 2x = 18 (AA), belonging to Panicoid, characterized by a comparatively small genome size of 515Mb approx (Lata *et al*, 2013). It belongs to the Panicoideae subfamily and the tribe Paniceae. The wild ancestor of cultivated foxtail millet (*S. italica*) is the green foxtail millet, *S. viridis* (2n = 2x = 18, AA). Wild species *S. faberii* and *S. verticillata* consist of an AABB genome, thought to have arisen from natural crossing between *S. viridis* and *S. adhaerans*. A species from Mexico, *S. grisebachii*, is a diploid species detected to have a CC genome. The sole species that is autotetraploid in the *Setaria* genus with an AAAA genome was identified in *S. queenslandica*. Other polyploid species identified in *Setaria* are *S. pumila* and *S. pallidefusca* (Benabdelmouna *et al*, 2001a; 2001b; Benabdelmouna and Darmency, 2003).

Genetic diversity

Enhancing crop resilience, production, and adaptation in adverse environmental conditions is largely dependent on genetic diversity. Establishing breeding goals and comprehending the genetic diversity of foxtail millet are important initial steps in developing superior cultivars that can meet the growing global demand (Ramesh *et al*, 2023). A wide range of genetic variation in the population of foxtail millet has been reported based on traits such as the number of tillers, bristle length, panicle orientation, panicle compactness, anther colour, and seed size (Wang *et al*, 2012; Moharil *et al*, 2019). ICRISAT analyzed approximately 1,535 *S. italica* samples from 26 different countries to look at differences in height of plant, time of blooming, structure of inflorescence, and the shape of seed amongst the collection. The possibility of gene transfer between Setaria species appears to be significant, especially between cultivated and wild species, since the outcrossing rate was shown to range significantly (0.3–4%). The variation in the crop has been explored through several studies, including pedigree, morphological and biochemical assessment within foxtail millet (Murugan and Nirmalakumari, 2006; Nirmalakumari and Vetriventhan, 2010). The biochemical analyses done in foxtail millet were based on seed protein analysis, isozymes (Jusuf and Pernes, 1985), and analysis of molecular markers or DNA markers (Schontz and Rether, 1999; Van *et al*, 2008; Fukunaga *et al*, 2002; Jia *et al*, 2009; Radha *et al*, 2014). Foxtail millet's genetic diversity is protected and encouraged, which might strengthen our ability to assist international efforts to develop a stronger and fairer food system and to encourage agricultural innovation (Govindaraj *et al*, 2020).

Breeding approaches in foxtail millet

Improving grain yield, nutritional composition, drought tolerance and resistance to pests and diseases while adapting to evolving demands can be achieved through the strategic use of diverse germplasm in breeding programmes. Below is an overview of breeding approaches employed in foxtail millet improvement (Figure 2).

Conventional approaches

In the early stages of foxtail millet breeding, pure line selection was the predominant method for enhancing grain yield. Breeding approaches for foxtail millet encompass selection and hybridization utilizing male-sterile lines (Swamy, 2023). Genetically male sterile lines governed by dominant gene 'Ch A' (Hu *et al*, 1986) and photoperiod-sensitive male sterility (Wensheng *et al*, 1991) have been developed for heterosis breeding in foxtail millet to facilitate the development of hybrid varieties (Liu *et al*, 2014). Hybridizing foxtail millet (*Setaria italica*) with its wild counterpart *S. viridis* was reported to incorporate triazine resistance into the cultivated species (Darmency and Pernes, 1985). *S. viridis* acts as an important genetic reservoir for improving foxtail millet, offering a straightforward and efficient breeding strategy (Naciri *et al*, 1992). The main challenge in conventional breeding is inbreeding depression, which results in the loss of heterosis or hybrid vigour. Another issue is linkage drag, where undesirable genes are transferred and integrated into the genome.

Mutation breeding approaches

Mutation breeding has significantly contributed to crop enhancement on a global scale. Its primary goal is to amplify the frequency and range of mutations while enhancing the occurrence of viable mutations for targeted genetic modifications. Induced mutagenesis acts as a viable alternative breeding strategy to enhance variability and address specific shortcomings in existing cultivars. Various mutant varieties were listed in the MVD (Mutant Variety Database), among which a few were foxtail millet (*S. italica* L.) resulting from direct or indirect mutation breeding with chemical and

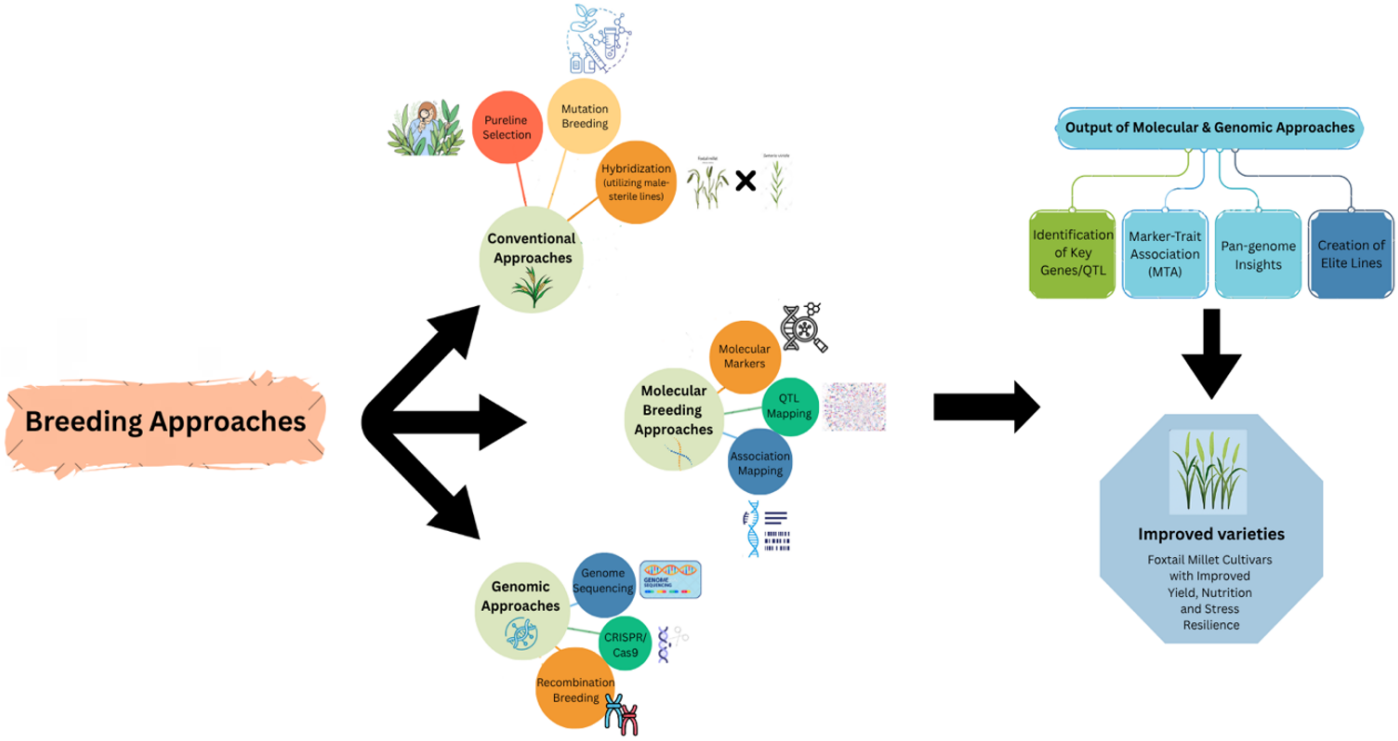


Figure 2. Overview of breeding approaches in foxtail millet. Created by the authors.

physical mutagens (IAEA, 2022). Some examples of mutant varieties released were Lugu 7, developed by irradiation of seeds with gamma rays with improved attributes of 10–15cm shorter stem and resistance to lodging, Jingu 47, officially approved in 2009, developed by treatment in aerospace, with main improved attributes of high yield and good quality. Also, mutagenesis techniques have been effectively combined with modern molecular biology tools, such as molecular marker analysis and high-throughput mutation screening, thereby increasing their efficiency and impact on crop improvement (Shu, 2009). Some examples in foxtail millet are SiDWARF2, a dwarf mutant gene (Xue *et al*, 2016), SiYGL1, a yellow-green leaf mutant (Li *et al*, 2016), and *siago1b*, a gene showing pleiotropic developmental defects (Liu *et al*, 2016).

Molecular breeding approaches

Conventional breeding often requires a long duration to develop and commercialize the cultivars, and several traits are easily influenced by environmental factors or exhibit low heritability. Advances in molecular biology led to the development of molecular marker technologies, which enable the study of plants based on polymorphic DNA sequences. Molecular markers are easily identified DNA fragments widespread in the genome with no environmental effect, and non-specific to tissue and stages of plants. It provides the most appropriate tool for the genetic diversity assessment, allowing the selection of suitable parental lines for further breeding programmes, efficient handling of plant genetic resources, and identifying varieties (de Vienne, 2003). The molecular approaches in foxtail millet are discussed broadly in the subsequent sections.

Molecular marker technologies

Molecular marker technologies have significantly advanced genetic studies and breeding programmes in foxtail millet, enabling precise identification of genetic variations. These markers are essential for assessing the genetic diversity, trait inheritance and crop improvement using marker-assisted selection (MAS). Among the molecular markers, restriction fragment length polymorphisms (RFLP) were the first molecular markers used in foxtail millet to analyze genetic differentiation across geographical diversity based on heterologous ribosomal DNA probes, providing foundational insights into domestication and diversity of foxtail millet and its wild ancestor *S. viridis* (Schontz and Rether, 1999; Fukunaga *et al*, 2002). RFLP map in the crop was constructed from an inter-varietal cross, Longgu 25 × Pagoda Flower green and reported nine linkage groups by Wang *et al* (1998). Doust *et al* (2004) expanded the work by using 257 RFLP markers from rice and foxtail millet to identify quantitative trait loci (QTL) in the aspect of branching and inflorescence architecture. Using F_2 plants derived from B100 (*S. italica*) and A10 (*S. viridis*) as a mapping population, the first functional dissection of morphological characters in foxtail millet using molecular mapping was developed. PCR-based molecular markers such as amplified fragment length polymorphism (AFLP), random amplified polymorphic DNA (RAPD), ISSR markers, and simple sequence repeats (SSRs) are widely used for genetic diversity studies, germplasm characterization, and linkage map construction (Gupta and Varshney, 2004; Jain *et al*, 2009; Kumari *et al*, 2011; Ardie *et al*, 2017). Among the molecular markers, SSR markers have been utilized to a great extent due to their reproducibility, high polymorphism and ease of use. This marker has been most extensively used in foxtail millet facilitating the study of diversity (Chander *et al*, 2017; Ramesh *et al*, 2023; Reddy *et*

al, 2025) for linkage map construction (Jia *et al*, 2009; Sato *et al*, 2013; Fang *et al*, 2016) and QTL mapping (Wang *et al*, 2017; Gao *et al*, 2025). The development of SSR markers based on the genomic resources of the draft genome in foxtail millet enables germplasm characterization, linkage mapping, phylogenetics and comparative mapping in foxtail millet (Pandey *et al*, 2013). Transposon display (TD marker) has been employed to examine the genetic structure of foxtail millet and wild green foxtail (*S. italica* subsp. *viridis* (L.) P. Beauv.) revealing geographical structuring (Hirano *et al*, 2011). The advent of single nucleotide polymorphism (SNP) has emerged as a powerful, high-resolution tool in molecular breeding and genomic research. The resequencing-based marker systems offers to detect SNPs and InDels and rapidly transitioned from low throughput to high throughput-based techniques and offers the construction of high-density linkage maps (Wang *et al*, 2017; Guo *et al*, 2023), genome wide association study (Jia *et al*, 2013; Jaiswal *et al*, 2019; 2024) and fine mapping of QTLs (Du *et al*, 2021; Li *et al*, 2022) in foxtail millet.

Methods for identification of QTLs in foxtail millet

Numerous chromosomal regions and QTLs have been identified that regulate a wide range of phenotypic traits in foxtail millet. According to Yu and Buckler (2006), the majority of the desirable features in plant breeding, such as height, quality, resistance to diseases, drought and salinity, are inherently quantitative in nature and are controlled by several genes. Such polygenic traits are regulated by specific genomic regions referred to as quantitative trait loci (QTLs). Phenotypic variations in complex traits arise from the combined effects of environmental factors, multiple QTL influences, QTL-QTL interactions, or QTL–environment interactions (Maloof, 2003). The primary goal of QTL identification is to locate genomic regions associated with complex traits and to identify tightly linked and neutral inherited molecular markers that can be used in breeding programmes (Agarwal *et al*, 2008). QTL identification involves the sequential arrangement of markers and determining genetic distances, assigning linkage groups based on recombination values. In order to address this complexity, high-resolution and accurate tools are required not only to decipher trait architecture but also to aid in the development of MAS tools for breeding programmes. QTL identification in plants is predominantly carried out using two approaches: (1) QTL mapping and (2) association mapping.

(1) QTL mapping in foxtail millet: Identified and genomic regions

The concept of QTL mapping was first introduced by Karl Sax in 1923. QTL mapping involves the analysis of an appropriate mapping population. Several QTL accounting for agronomic and yield attribute traits were reported in foxtail millet (Doust *et al*, 2004; Wang *et al*, 2013; Fang *et al*, 2016). QTL mapping was employed in the crop for *stb1* gene, for 'spikelet-tipped bristles' (Sato *et al*, 2013), variations in blooming time under various environmental circumstances (Mauro-Herrera *et al*, 2013), and germination and drought tolerance in early planting (Qie *et al*, 2014). The integration of next-generation sequencing enabled the development of high-density SNP based linkage maps. Map-based cloning was used to fine map and clone SiDWARF2 (a foxtail millet dwarf mutant) derived from Yugu1 (Xue *et al*, 2016), a recessive nuclear gene SiYGL1 encoding a magnesium-chelatase D subunit, responsible for yellow–green leaf mutant (Li *et al*, 2016), and Argonaute 1 (AGO1) mutant (*siago1b*) induced by ethyl methane sulfonate exhibiting pleiotropic developmental defects in foxtail millet (Liu *et al*, 2016). Liu *et al* (2024) identified QTLs associated with downy mildew resistance based on the specific locus amplified fragment sequencing with high-density linkage map and the phenotype data in four environments. They also concluded through collinearity analysis between genomes of pearl millet and foxtail millet that the genes were taxon-specific (Liu *et al*, 2024). Ma *et al* (2025) identified four important hull colour QTL and provided a basis for characterizing hull colour indices and contributing to the advancement of QTL mapping for grain colour. Liu *et al* (2022) identified 221 QTLs for 17 morpho-agronomic and yield-related traits, while Yoshitsu *et al* (2017) identified two QTLs (qDTH2 and qDTH7) regulating days to heading (DTH) using QTL-seq analysis. Ni *et al* (2017) performed resequencing of 184 recombinant inbred lines (RILs) of foxtail millet for QTL mapping of nine agronomic traits and identified a single gene controlling five traits, two QTLs for plant height and three QTLs for heading date. This provides an efficient way for constructing a high-resolution genome assembly and identifying genes. Han *et al* (2024) developed a genetic map for plant height based on re-sequencing, identifying 19 unconditional and 13 conditional QTLs. QTLs identified using molecular markers in foxtail millet are shown in Table 1.

(2) Advances in association mapping or association analysis in foxtail millet

Association mapping or association analysis uses linkage disequilibrium (LD) to investigate the connection between genotypic constitution and phenotypic expression in natural populations (Borba *et al*, 2010). It is typically conducted using two approaches – candidate gene-based and genome-wide association mapping. The major advantages of this approach are its use of natural populations, the elimination of bi-parental population requirements, and its ability to achieve high-resolution mapping due to recombination events accumulated over multiple generations (Agrama *et al*, 2007). GWAS was performed using the core collection of foxtail millet as an association mapping panel, which involves genome-wide screening of nucleotide sequence variation. Gupta *et al* 2014, evaluated 50 SST markers on 184 foxtail millet accessions across nine chromosomes and identified eight significant markers linked to agronomic traits and two markers with significant association for ubiquitin carboxyl-terminal hydrolase and phospholipid acyltransferase. GWAS by Upadhyaya *et al* (2015) identified several SNP loci associated, as well as a major genomic region of plant pigmentation and flowering time. The genomic regions of plant pigmentation were identified between 7.2 and 7.3Mbp on chromosome 4, which was also reported by Jia *et al* (2013) for the pigmentation-related traits such as colour of bristle, leaf sheath and pulvinus. Significant associations were identified on chromosome 6, specifically around 34.0 to 35.5Mbp.

Table 1. A summary of QTL identification in foxtail millet. Abbreviations: QTL, quantitative trait loci; RFLP, restriction fragment length polymorphism; SSR, simple sequence repeat; GWAS, genome-wide association study; SNP, single nucleotide polymorphism; GBS, genotyping-by-sequencing; RAD-seq, restriction site-associated DNA sequencing; MTA, marker–trait association; RIL, recombinant inbred line; BSR-seq, bulked segregant RNA sequencing; SLAF-seq, specific length amplified fragment sequencing; InDel, insertion–deletion; ICIM, inclusive composite interval mapping; MCIM, mixed composite interval mapping.

Germplasm	Mapping approach	Population type	Marker used	No. of markers	Genomic region(s) identified	Putative gene identified	Trait(s) studied	References
Cross between *Setaria italica* (B100) × *S. viridis* (A10)	QTL mapping (Composite Interval Mapping, comparative genomics)	F2:3 population (~120 families)	RFLP markers	119 markers (subset from 257 loci map)	Multiple QTL across chromosomes I–IX; major regions on chromosomes V, VI, IX	tb1 (minor effect), auxin & gibberellin pathway genes, monoculm1, dwarf-related genes	Tillering, axillary branching	Doust *et al*, 2004
Diverse germplasm panel	Association mapping	Diverse germplasm panel, 184 foxtail millet accessions	Genomic SSRs	50	Strong associations on chromosome 5	*SSR b129* (Ubiquitin carboxyl-terminal hydrolase) and SSR *p75* (Phospholipid acyltransferase)	Yield-related agronomic attributes	Gupta *et al*, 2014
Global Collection	GWAS Association Mapping)	181 accessions, including 155 from the core collection	GBS	17,714	Pigmentation and days to 50% flowering at chromosomes 4 and 6, respectively	-	Flowering time and plant pigmentation	Upadhyaya *et al*, 2015
Segregating population	Linkage mapping	F_2 population (Yugu1 × Longgu7, 167 individuals)	SSRs	10,598	QTLs distributed across all nine chromosomes	-	Agronomic and yield attributes	Fang *et al*, 2016
RILs	High-density linkage mapping and QTL analysis	184 RILs (Zhanggu × A2)	SNPs	483,414 SNPs; 3,437 recombination bins	QTLs on all nine chromosomes	*sd1* (gibberellin synthesis gene)	Agronomic traits	Ni *et al*, 2017
Segregating population	RAD-seq based QTL mapping	F_2 population from cross between Hongmiaozhangu × Changnong35 (124 individuals)	SNPs	9,968	11 QTLs on chromosome 1,2,5,7,8,9	-	Agronomic traits	Wang *et al*, 2017
Yuikogane × Shinanotsubuhime (F_2 population, 382 plants)	QTL-seq combined with Bulked Segregant Analysis (BSA) and Composite Interval Mapping (CIM)	F_2 population	SNPs, Indel markers, CAPS markers	~45,370 SNPs identified between parents; 24 Indel & CAPS markers used for validation	Two major QTLs: qDTH2 (Chr 2: 38.2–39.6Mb) and qDTH7 (Chr 7: 29.2–31.0Mb)	Candidate genes include Seita.2G286100 (OsPRR95 homolog), Seita.2G291300 (DLF1 homolog) and Seita.7G246700 (Roc4 homolog)	Days to Heading (DTH) / Flowering time	Yoshitsu *et al*, 2017
Cultivars	GWAS Association Mapping	142 diverse genotypes (core collection)	SNPs (GBS-ddRAD)	12,460	High-confidence MTAs on chromosomes 3, 6, 7, 9	27 candidate genes for 1,000-grain weight, grain yield and flag leaf width	Major agronomic traits	Jaiswal *et al*, 2019

Table 1 continued

Germplasm	Mapping approach	Population type	Marker used	No. of markers	Genomic region(s) identified	Putative gene identified	Trait(s) studied	References
Yugu1 × Longgu7	QTL mapping (linkage mapping using resequencing data)	RIL population	Bin markers + SSR markers	2297 bin markers + 74 SSR markers	221 QTL across genome; 22 QTL clusters; stable QTL qLMS6.1 (Chr 6)	Seita.6G250500	Morpho-agronomic and yield-related traits (plant height, tiller number, panicle traits, yield traits)	Liu *et al*, 2022
Aininghuang × Jingu21	QTL mapping (unconditional & conditional QTL analysis + transcriptome integration)	RIL population	Bin markers (resequencing-based)	4360 bin markers	19 unconditional QTL + 13 conditional QTL; 4 stable QTL across environments	8 candidate genes (identified via RNA-seq & WGCNA; specific names not mentioned in abstract)	Plant height (PH)	Han *et al*, 2024
Foxtail millet G1 × JG21	QTL mapping using high-density linkage map and BSR-seq validation	Recombinant Inbred Lines (F6:7, 158 lines)	SLAF-seq SNP bin markers	1031 bin markers	Major region on Chr8 (0.78 Mb interval) including qDM8_1, qDM8_2, qDM8_4	Seita8G.199800, Seita8G.195900, Seita8G.198300, Seita8G.199300 (NBS-LRR genes)	Downy mildew resistance	Liu *et al*, 2024
Foxtail millet Jingu28 × Ai88	QTL mapping using genetic linkage map	F_2 population (300 individuals)	SSR + InDel markers	215 (213 SSR + 2 InDel)	Major QTLs qHD9-1 (Chr9) and qPH5-1 (Chr5); 46 QTLs forming 13 clusters	Seita.9G020100 (CCT motif gene), Seita.5G404900 (GA20-oxidase)	12 agronomic traits including heading date and plant height	Gao *et al*, 2025
Foxtail millet Changsheng07 × Donggu218	QTL mapping using genetic linkage map	F_2 population with $F_{2:3}$ families	SSR + InDel markers	196 (159 SSR + 37 InDel)	Major QTLs qMPL3.1, qMPL5, qMPW2, qSD5, qTGW5.1, qTGW5.2, qGL5; 22 QTLs forming 4 clusters	-	Panicle-related traits (panicle length, width, grain length, TGW)	Li *et al*, 2025
Yugu18 × Hongjiugu19	QTL mapping (ICIM, MCIM, multi-method phenotyping)	RIL (F6, 250 lines)	SNPs, InDels, bin markers	~20,748 SNPs + 1,759 InDels (1420 bins)	Major QTL: qHC1.1, qHC1.2 (Chr 1); qHC9.1, qHC9.3 (Chr 9)	Not specifically identified (QTL regions validated; overlap with previous loci)	Hull colour (grain colour traits using 4 methods)	Ma *et al*, 2025

Advances in genomic research for foxtail millet improvement

Among minor millets, foxtail millet possesses the smallest genome (423–510Mb) and was the first millet crop to have its entire genome sequenced. Its compact diploid genome, rapid growth cycle and self-pollination make it a model for C4 species (Vetriventhan *et al*, 2020). Zhang *et al* (2012) made a major breakthrough, producing a draft genome (~423Mb) anchored to nine chromosomes. This provides marker discovery, gene annotation and comparative genomics with other cereals such as rice, pearl millet and sorghum. The reference genome allowed the development of a genome-wide scale of microsatellite markers across the nine chromosomes of foxtail millet. A diverse foxtail millet collection was genotyped through genotyping-by-sequencing (GBS) by ICRISAT in collaboration with Cornell University, identifying genome-wide SNPs and assessing population structure and diversity (Upadhyaya *et al*, 2015). Several gene families have been identified and characterised in foxtail millet, including NAC (NAM, ATAF, and CUC) (Puranic *et al*, 2013), MYB (myeloblastosis transcription factor) (Muthamilarasan *et al*, 2014b), WRKY (WRKY DNA-binding protein) (Muthamilarasan *et al*, 2015), AP2/ERF (APETALA2/Ethylene-Responsive Factor) (Lata *et al*, 2014), and C2H2 (Cys2-His2 zinc finger protein) (Muthamilarasan *et al*, 2014a), which play important roles in stress response, growth, and development. In addition, regulatory components such as DCL (Dicer-like), AGO (Argonaute), and RDR (RNA-dependent RNA polymerase) (Yadav *et al*, 2015) are involved in gene silencing pathways. Although these gene families are conserved across plant species, their characterisation in foxtail millet provides insights into mechanisms of stress tolerance and can aid genetic improvement.

The Beijing Genomics Institute, China and The Joint Genome Institute, USA, have recently sequenced the complete genome of two foxtail millet accessions. The comparative genome mapping through sequence alignment of foxtail millet demonstrated a strong syntenic relationship with both rice and sorghum, even though they have diverged over half a century years ago. With this genome-wide sequence resources and availability in public repositories, it is now possible to develop *in silico* molecular markers and large-scale validation to utilize them in various applications for genetic improvement in foxtail millet. Such a genome-wide scale may prove beneficial for other underutilized or orphan crop species with limited or no genomic information available.

Application of genome editing technologies in foxtail millet

Genome editing technologies have developed as a powerful tool because of their precision and simplicity in modifying targeted genomic regions by using engineered nucleases such as zinc finger nucleases (ZFNs), transcription activator-like effector nucleases (TAL-ENs), or RNA-guided engineered nucleases such as the CRISPR/Cas system. The first genome editing in foxtail millet was successfully demonstrated by Lin *et al* (2018) through mutating the *SiPDS* gene, which was achieved through protoplast transfection. Cheng *et al* (2021) targeted gene *SiMTL* by using CRISPR/Cas9 to generate a haploid inducer line. Although there are relevant reports of genome editing systems in foxtail millet, a highly efficient genome-editing system is limited. Liang *et al* (2022) first successfully applied base editing in foxtail millet by using CRISPR/Cas9 targeting the *SiALS* gene, and successfully created a homozygous mutant plant that is herbicide-tolerant.

Future directions and challenges

Foxtail millet is a significant crop with its climatic resilience and superior nutritional profile. With a potential role in food and nutritional security, it is necessary to harness its nutritional potential to meet daily needs for a healthy lifestyle. Addressing the challenges for breeding foxtail millets is necessary to accomplish this goal. An in-depth understanding of the floral biology of foxtail millet is necessary for developing standardized hybridization techniques to create variability, develop improved varieties and enhance tolerance to adverse climatic conditions, which shall serve the breeding efforts of the crop. Exploration of underutilized germplasm, including wild relatives, shall also enhance biotic and abiotic stresses and nutritional quality. The use of advanced genomic tools (e.g. marker-assisted selection and genome-wide association studies) has improved foxtail millet breeding. Collaboration among genebanks, researchers and policymakers is crucial for sustainable utilization of foxtail millet genetic resources (Vetriventhan *et al*, 2020).

Conclusion

Foxtail millet holds immense potential as a climate-resilient, nutrient-rich crop that can address global food security challenges. Advances in breeding methodologies, including molecular marker technologies and genomic selection, have enabled precise identification and introgression of desirable traits. QTL mapping and association studies have identified key genetic loci governing yield, stress tolerance, and disease resistance, laying the foundation for marker-assisted breeding. The availability of high-throughput sequencing and genome-editing tools such as CRISPR/Cas9 has further enhanced the scope of foxtail millet improvement. However, challenges such as limited genetic diversity in cultivated varieties, underutilization of wild germplasm, and the need for better characterization of genomic resources persist. Further research should prioritize integrating advanced genomic tools, promoting international germplasm exchange and developing climate-resilient cultivars. By leveraging genetic innovations and sustainable breeding approaches, foxtail millet can play an important role in uplifting global agricultural productivity and ensuring food security amid climate change.

Author contributions

Shivika Pareek: Compilation of the literature, preparation of the draft and preparation of figures/tables; Reginah Pheirim: Conceptualization and article revision; Chetariya Chana Pitha: Revision of article; Alka Soharu: Revision of article.

Conflict of interest statement

The authors have no relevant financial or non-financial competing interests to disclose.

References

Agarwal, M., Shrivastava, N., Padh, H. (2008). Advances in molecular marker techniques and their applications in plant sciences. *Plant Cell Reports*, 27(4), 617-631. https://doi.org/10.1007/s00299-008-0507-z

Agrama, H. A., Eizenga, G. C., Yan, W. (2007). Association mapping of yield and its components in rice cultivars. *Molecular Breeding*, *19*(4), 341-356. https://doi.org/10.1007/s11032-006-9066-6

Ardie, S., Khumaida, N., Fauziah, N., Yudiansyah, Y. (2017). Biodiversity assessment of foxtail millet (*Setaria italica* L.) genotypes based on RAPD marker. *J. Trop. Crop Sci.* 4:21–25. https://doi.org/10.29244/jtcs.4.1.21-25

Benabdelmouna, A., Darmency, H. (2003). Copia-like retrotransposons in the genus *Setaria*: Sequence heterogeneity, species distribution and chromosomal organization. *Plant Systematics and Evolution*, *237*(3), 127-136. https://doi.org/10.1007/s00606-002-0250-9

Benabdelmouna, A., Abirached-Darmency, M., Darmency, H. (2001a). Phylogenetic and genomic relationships in Setaria italica and its close relatives based on the molecular diversity and chromosomal organization of 5S and 18S-5.8 S-25S rDNA genes. *Theoretical and Applied Genetics*, *103*(5), 668-677. https://doi.org/10.1007/s001220100596

Benabdelmouna, A., Shi, Y., Abirached-Darmency, M., & Darmency, H. (2001b). Genomic in situ hybridization (GISH) discriminates between the A and the B genomes in diploid and tetraploid *Setaria* species. *Genome*, *44*(4), 685-690. https://doi.org/10.1139/g01-032

Borba, T. C. D. O., Brondani, R. P. V., Breseghello, F., Coelho, A. S. G., Mendonça, J. A., Rangel, P. H. N., Brondani, C. (2010). Association mapping for yield and grain quality traits in rice (*Oryza sativa* L.). *Genetics and Molecular Biology*, *33*, 515-524. https://doi.org/10.1590/S1415-47572010005000065

Bramel, P., Giovannini, P., Eshan Dulloo, M. (2022). Global strategy for the conservation and use of genetic resources of selected millets. *Global Crop Diversity Trust, Bonn*, *85*.

Brink, M. (2006). *Setaria italica* (L.) P. Beauv. *Cereals and Pulses, Plant Resources of Tropical Africa*.

Chander, S., Bhat, K. V., Kumari, R., Sen, S., Gaikwad, A. B., Gowda, M. V. C., Dikshit, N. (2017). Analysis of spatial distribution of genetic diversity and validation of Indian foxtail millet core collection. *Physiology and Molecular Biology of Plants*, 23(3), 663-673. https://doi.org/10.1007/s12298-017-0448-5

Cheng, Z., Sun, Y., Yang, S., Zhi, H., Yin, T., Ma, X., Sui, Y. (2021). Establishing in planta haploid inducer line by edited SiMTL in foxtail millet (*Setaria italica*). Plant Biotechnology Journal, 19(6), 1089. https://doi.org/10.1111/pbi.13584

Darmency, H., Pernes, J. (1985). Use of wild *Setaria viridis* (L.) Beauv. to improve triazine resistance in cultivated *S. italica* (L.) by hybridization. *Weed Research*, *25*(3), 175-179. https://doi.org/10.1111/j.1365-3180.1985.tb00633.x

de Vienne, D. (Ed.). (2003). *Molecular markers in plant genetics and biotechnology*. CRC Press.

Dekker, J. (2003). The foxtail (*Setaria*) species-group. *Weed science*, *51*(5), 641-656. https://doi.org/10.1614/P2002-IR

Doust, A. N., Devos, K. M., Gadberry, M. D., Gale, M. D., Kellogg, E. A. (2004). Genetic control of branching in foxtail millet. *Proceedings of the National Academy of Sciences*, *101*(24), 9045-9050. https://doi.org/10.1073/pnas.0402892101

Du, X., Wang, Z., Han, K., Lian, S., Li, Y., Zhang, L., Guo, E., Wang, J. (2021). Fine mapping of qPH9, a major quantitative trait locus, responsible for plant height in foxtail millet [*Setaria italica* (L.) P. Beauv.]. *Molecular Breeding*, *41*(12), 77. https://doi.org/10.1007/s11032-021-01261-w

Dwivedi, S. L., Upadhyaya, H. D., Senthilvel, S., Hash, C. T., Fukunaga, K., Diao, X., Prasad, M. (2012). Millets: genetic and genomic resources. *Plant Breed Rev.* 2012; 35:247–375.

Fang, X., Dong, K., Wang, X., Liu, T., He, J., Ren, R., Zhang, L., Liu, R., Liu, X., Li, M., Huang, M., Zhang, Z., Yang, T. (2016). A high-density genetic map and QTL for agronomic and yield traits in foxtail millet [*Setaria italica* (L.) P. Beauv.]. *BMC Genomics* 17:336. https://doi.org/10.1186/s12864-016-2628-z

Fukunaga, K., Kawase, M. (2024). Crop evolution of foxtail millet. *Plants*, *13*(2), 218. https://doi.org/10.3390/plants13020218

Fukunaga, K., Wang, Z., Kato, K., Kawase, M. (2002). Geographical variation of nuclear genome RFLPs and genetic differentiation in foxtail millet, *Setaria italica* (L.) P. Beauv. *Genetic Resources and Crop Evolution*, 49(1), 95-101. https://doi.org/10.1023/A:1013852007770

Gao, L., Zhu, Q., Li, H., Wang, S., Fan, J., Wang, T., Yang, G. (2025). Construction of a genetic linkage map and QTL mapping of the agronomic traits in Foxtail millet (*Setaria italica*). *BMC Genomics*, 26(1), 152. https://doi.org/10.1186/s12864-024-11169-2

Govindaraj, M., Rai, K. N., Kanatti, A., Upadhyaya, H. D., Shivade, H., & Rao, A. S. (2020). Exploring the genetic variability and diversity of pearl millet core collection germplasm for grain nutritional traits improvement. *Scientific Reports*, *10*(1), 21177. https://doi.org/10.1038/s41598-020-77818-0

Guo, S., Chai, S., Guo, Y., Shi, X., Han, F., Qu, T., ... & Yang, P. (2023). Mapping of major QTL and candidate gene analysis for hull colour in foxtail millet (*Setaria italica* (L.) P. Beauv.). *BMC Genomics*, *24*(1), 458. https://doi.org/10.1186/s12864-023-09517-9

Gupta, P. K., Varshney, R. K. (2004). Cereal genomics: an overview. *Cereal genomics*, 1-18.

Gupta, S., Kumari, K., Muthamilarasan, M., Parida, S. K., Prasad, M. (2014). Population structure and association mapping of yield contributing agronomic traits in foxtail millet. *Plant Cell Reports*, *33*(6), 881-893. https://doi.org/10.1007/s00299-014-1564-0

Han K, Wang Z, Shen L, Du X, Lian S, Li Y, Li Y, Tang C, Li H, Zhang L, Wang J (2024) Mapping of dynamic quantitative trait loci for plant height in a RIL population of foxtail millet (*Setaria italica* L.). *Frontiers in Plant Science*, *15*, 1418328. https://doi.org/10.3389/fpls.2024.1418328

Hirano, R., Naito, K., Fukunaga, K., Watanabe, K. N., Ohsawa, R., Kawase, M. (2011). Genetic structure of landraces in foxtail millet (*Setaria italica* (L.) P. Beauv.) revealed with transposon display and interpretation to crop evolution of foxtail millet. *Genome*, 54(6), 498-506. https://doi.org/10.1139/g11-015

Hu, H. K., Ma, S. Y., Shi, Y. H. (1986). The discovery of a dominant male-sterile gene in millet (*Setaria italica*). *Acta Agronomica Sinica* 12, 73-78

IAEA (2022). Mutant Variety Database. https://mvd.iaea.

org/
Jain, S. M., Brar, D. S., Ahloowalia, B. S. (Eds.). (2009). Molecular techniques in crop improvement. Dordrecht: Springer Netherlands.
Jaiswal, V., Bandyopadhyay, T., Singh, R. K., Gahlaut, V., Muthamilarasan, M., Prasad, M. (2024). Multi-environment GWAS identifies genomic regions underlying grain nutrient traits in foxtail millet (*Setaria italica*). *Plant Cell Reports*, 43(1), 6. https://doi.org/10.1007/s00299-023-03127-1
Jaiswal, V., Gupta, S., Gahlaut, V., Muthamilarasan, M., Bandyopadhyay, T., Ramchiary, N., Prasad, M. (2019). Genome-wide association study of major agronomic traits in foxtail millet (*Setaria italica* L.) using ddRAD sequencing. *Scientific Reports*, 9(1), 5020. https://doi.org/10.1038/s41598-019-41602-6
Jayaraman N, Suresh S, Nirmala A, Ganeshan NM (1997) Genetic enhancement and breeding strategies in small millets. In: National Seminar on Small Millets, 23-24 April, 1997, Coimbatore, India, pp 19-21 (Extended summaries)
Jia, G., Huang, X., Zhi, H., Zhao, Y., Zhao, Q., Li, W., Han, B. (2013). A haplotype map of genomic variations and genome-wide association studies of agronomic traits in foxtail millet (*Setaria italica*). *Nature Genetics*, 45(8), 957-961. https://doi.org/10.1038/ng.2673
Jia, X., Zhang, Z., Liu, Y., Zhang, C., Shi, Y., Song, Y., Li, Y. (2009). Development and genetic mapping of SSR markers in foxtail millet [*Setaria italica* (L.) P. Beauv.]. *Theoretical and Applied Genetics*, 118(4), 821-829. https://doi.org/10.1007/s00122-008-0942-9.
Jusuf, M., Pernes, J. (1985). Genetic variability of foxtail millet (*Setaria italica* P. Beauv.) Electrophoretic study of five isoenzyme systems. *Theoretical and Applied Genetics*, 71(3), 385-391. https://doi.org/10.1007/BF00251177
Kumari, R., Dikshit, N., Sharma, D., Bhat, K. V. (2011). Analysis of molecular genetic diversity in a representative collection of foxtail millet (*Setaria italica* (L.) P. Beauv.) from different agro-ecological regions of India. *Physiology and Molecular Biology of Plants*, 17(4), 363–374. https://doi.org/10.1007/s12298-011-0085-3
Lata, C., Gupta, S., Prasad, M. (2013). Foxtail millet: a model crop for genetic and genomic studies in bioenergy grasses. *Critical reviews in biotechnology*, 33(3), 328-343. https://doi.org/10.3109/07388551.2012.716809
Lata, C., Mishra, A. K., Muthamilarasan, M., Bonthala, V. S., Khan, Y., Prasad, M. (2014). Genome-wide investigation and expression profiling of AP2/ERF transcription factor superfamily in foxtail millet (*Setaria italica* L.). *PLoS One*, 9(11), e113092. https://doi.org/10.1371/journal.pone.0113092
Li, W., Tang, S., Zhang, S., Shan, J., Tang, C., Chen, Q., Diao, X. (2016). Gene mapping and functional analysis of the novel leaf color gene SiYGL1 in foxtail millet [*Setaria italica* (L.) P. Beauv]. *Physiologia Plantarum*, 157(1), 24-37. https://doi.org/10.1111/ppl.12405
Li, Y., Wu, S., Cao, Y. (1995). Cluster analysis of an international collection of foxtail millet (*Setaria italica* (L.) P. Beauv.). *Euphytica*, 83(1), 79-85. https://doi.org/10.1007/BF01677864
Li, F., Niu, J., Yu, X., Kong, Q., Wang, R., Qin, L., Chen, E., Yang, Y., Liu, Z., Lang, L., Zhang, H., Wang, H., Guan, Y. (2022). Isolation and identification of SiCOL5, which is involved in photoperiod response, based on the quantitative trait locus mapping of *Setaria italica*. *Frontiers in Plant Science*, 13, 969604. https://doi.org/10.3389/fpls.2022.969604
Li, Y., Wu, S. (1996). Traditional maintenance and multiplication of foxtail millet (*Setaria italica* (L.) P. Beauv.) landraces in China. *Euphytica*, 87(1), 33-38. https://doi.org/10.1007/BF00022961
Li, H., Zhu, Q., Wang, T., Zhao, Y., Fan, J., Yang, L., Yang, G. (2025). Genetic linkage map construction and QTL mapping for panicle-related traits in foxtail millet (*Setaria Italica* (L.) P. Beauv.). *Molecular Breeding*, 45(9), 72. https://doi.org/10.21203/rs.3.rs-5367279/v1
Liang, Z., Wu, Y., Ma, L., Guo, Y., Ran, Y. (2022). Efficient genome editing in *Setaria italica* using CRISPR/Cas9 and base editors. *Frontiers in Plant Science*, 12, 815946. https://doi.org/10.3389/fpls.2021.815946
Lin, C. S., Hsu, C. T., Yang, L. H., Lee, L. Y., Fu, J. Y., Cheng, Q. W., Shih, M. C. (2018). Application of protoplast technology to CRISPR/Cas9 mutagenesis: from single-cell mutation detection to mutant plant regeneration. *Plant Biotechnology Journal*, 16(7), 1295-1310. https://doi.org/10.1111/pbi.12870
Liu, X., Tang, S., Jia, G., Schnable, J. C., Su, H., Tang, C., Diao, X. (2016). The C-terminal motif of SiAGO1b is required for the regulation of growth, development and stress responses in foxtail millet (*Setaria italica* (L.) P. Beauv). *Journal of Experimental Botany*, 67(11), 3237-3249. https://doi.org/10.1093/jxb/erw135
Liu, Z., Zhang, T., Li, C., Bai, G. (2014). Genetic diversity and classification of cytoplasm of Chinese elite foxtail millet [*Setaria italica* (L.) P. Beauv.] germplasm. *Crop Science*, 54(2), 659-666. https://doi.org/10.2135/cropsci2012.11.0646
Liu, T., He, J., Dong, K., Wang, X., Zhang, L., Ren, R., Zhang, Z. (2022). Genome-wide identification of quantitative trait loci for morpho-agronomic and yield-related traits in foxtail millet (*Setaria italica*) across multi-environments. *Molecular Genetics and Genomics*, 297(3), 873-888. https://doi.org/10.1007/s00438-022-01894-2
Liu, X., Zhang, N., Sun, Y., Fu, Z., Han, Y., Yang, Y., Jia, J., Hou, S., Zhang, B. (2024). QTL mapping of downy mildew resistance in foxtail millet by SLAF-seq and BSR-seq analysis. *Theoretical and Applied Genetics*, 137(7), 168. https://doi.org/10.1007/s00122-024-04673-2
Ma, Z., Chai, S., Wu, Y., Li, Y., Han, H., Song, H., Gao, J., Feng, B., Yang, P. (2025). QTL identification of hull color for foxtail millet [*Setaria italica* (L.) P. Beauv.] through four phenotype identification strategies in a RIL population. *Seeds*, 4(1), 10. https://doi.org/10.3390/seeds4010010
Maloof, J. N. (2003). QTL for plant growth and morphology. *Current opinion in plant biology*, 6(1), 85-90. https://doi.org/10.1016/S1369526602000080
Mauro-Herrera, M., Wang, X., Barbier, H., Brutnell, T. P., Devos, K. M., Doust, A. N. (2013). Genetic control and comparative genomic analysis of flowering time in *Setaria* (Poaceae). *G3: Genes| Genomes| Genetics,* 3(2), 283-295. https://doi.org/10.1534/g3.112.005207.
Moharil, M. P., Ingle, K. P., Jadhav, P. V., Gawai, D. C., Khelurkar, V. C., Suprasanna, P. (2019). Foxtail millet (*Setaria italica* L.): potential of smaller millet for future breeding. *Advances in Plant Breeding Strategies: Cereals:* Volume 5, 133-163. https://doi.org/10.1007/978-3-030-23108-8_4
Murugan R, Nirmalakumari A (2006) Genetic divergence in foxtail millet (*Setaria italica* (L.) Beauv.). *Indian J Genet*

66(4):339–340
Muthamilarasan, M., Bonthala, V. S., Khandelwal, R., Jaishankar, J., Shweta, S., Nawaz, K., Prasad, M. (2015). Global analysis of WRKY transcription factor superfamily in *Setaria* identifies potential candidates involved in abiotic stress signaling. *Frontiers in Plant Science*, 6, 910. https://doi.org/10.3389/fpls.2015.00910
Muthamilarasan, M., Bonthala, V. S., Mishra, A. K., Khandelwal, R., Khan, Y., Roy, R., Prasad, M. (2014a). C2H2 type of zinc finger transcription factors in foxtail millet define response to abiotic stresses. *Functional & Integrative Genomics*, 14(3), 531-543. https://doi.org/10.1007/s10142-014-0383-2
Muthamilarasan, M., Khandelwal, R., Yadav, C. B., Bonthala, V. S., Khan, Y., Prasad, M. (2014b). Identification and molecular characterization of MYB transcription factor superfamily in C4 model plant foxtail millet (*Setaria italica* L.). *PLoS One*, 9(10), e109920. https://doi.org/10.1371/journal.pone.0109920
Naciri, Y., Darmency, H., Belliard, J., Dessaint, F., Pernès, J. (1992). Breeding strategy in foxtail millet, *Setaria italica* (LP Beauv.), following interspecific hybridization. *Euphytica*, 60(2), 97-103. https://doi.org/10.1007/BF00029664
Ni, X., Xia, Q., Zhang, H., Quan, Z. (2017). Updated foxtail millet genome assembly and gene mapping of nine key agronomic traits by resequencing a RIL population. *GigaScience,* 6(2), 1–8. https://doi.org/10.1093/gigascience/giw005
Nirmalakumari, A., Vetriventhan, M. (2010). Characterization of foxtail millet germplasm collections for yield contributing traits. *Electronic Journal of Plant Breeding*, 1(2), 140-147.
Pandey, G., Misra, G., Kumari, K., Gupta, S., Parida, S. K., Chattopadhyay, D., Prasad, M. (2013). Genome-wide development and use of microsatellite markers for large-scale genotyping applications in foxtail millet (*Setaria italica* (L.)). *DNA Research*, 20(2), 197–207. https://doi.org/10.1093/dnares/dst002
Puranik, S., Sahu, P. P., Mandal, S. N., B, V. S., Parida, S. K., Prasad, M. (2013). Comprehensive genome-wide survey, genomic constitution and expression profiling of the NAC transcription factor family in foxtail millet (*Setaria italica* L.). *PloS One*, 8(5), e64594. https://doi.org/10.1371/journal.pone.0064594
Qie, L., Jia, G., Zhang, W., Schnable, J., Shang, Z., Li, W., Diao, X. (2014). Mapping of quantitative trait locus (QTLs) that contribute to germination and early seedling drought tolerance in the interspecific cross *Setaria italica*× *Setaria viridis*. *PloS One,* 9(7), e101868. https://doi.org/10.1371/journal.pone.0101868
Radha, B. N., Channakeshava, B. C., Bhanuprakash, K. (2014). Biochemical landmarks for Identification of foxtail millet genotypes. *Intl J Appl*, 9(1), 9-26.
Rahayu, M., Jansen, P. C. M. (1996). *Setaria italica* (L.) P. Beauvois cv. group foxtail millet. *Plant Resources of South-East Asia*, (10), 127-130.
Ramesh, P., Juturu, V. N., Yugandhar, P., Pedersen, S., Hemasundar, A., Yolcu, S., Chandra Sekhar, A. (2023). Molecular genetics and phenotypic assessment of foxtail millet (*Setaria italica* (L.) P. Beauv.) landraces revealed remarkable variability of morpho-physiological, yield, and yield-related traits. *Frontiers in Genetics*, 14, 1052575. https://doi.org/10.3389/fgene.2023.1052575
Reddy, M. P., Rai, V. P., Reddy, C. C. M., Kumar, R., Paliwal, R., Sinha, B. (2025). Genetic diversity and population structure analyses using hypervariable microsatellite markers in foxtail millets (*Setaria* species): Future smart nutri-cereal crop. *South African Journal of Botany*, 177, 329-337. https://doi.org/10.1016/j.sajb.2024.11.043
Sato, K., Mukainari, Y., Naito, K., Fukunaga, K. (2013). Construction of a foxtail millet linkage map and mapping of spikelet-tipped bristles 1 (stb1) by using transposon display markers and simple sequence repeat markers with genome sequence information. *Molecular Breeding*, 31(3), 675-684. https://doi.org/10.1007/s11032-012-9825-5
Schontz, D., Rether, B. (1999). Genetic variability in foxtail millet, *Setaria italica* (L.) P. Beauv.: identification and classification of lines with RAPD markers. *Plant Breeding*, 118(2), 190-192. https://doi.org/10.1046/j.1439-0523
Shu, Q. Y. (2009). A summary of the international symposium on induced mutations in plants. In: Shu QY (ed) Induced plant mutations in the genomics era. FAO, UN, Rome, pp 15–18.
Siles, M. M., Baltensperger, D. D., Nelson, L. A. (2001). Technique for artificial hybridization of foxtail millet [*Setaria italica* (L.) Beauv.]. *Crop Science*, 41(5), 1408-1412. https://doi.org/10.2135/cropsci2001.4151408x
Singh, D., Lawrence, K., Marker, S., Bhattacharjee, I., Lawrence, R., Choudhary, R., Karunakaran, R. (2023). Rainfed assessment of foxtail millet (*Setaria italica* L. beauv) germplasms through genotyping and principal component analysis. *Frontiers in Plant Science*, 14, 1017652. https://doi.org/10.3389/fpls.2023.1017652
Sundararaj DP, Thulasidas G (1976) Botany of Field Crops, Macmillan Publisher, India, 509 pp
Swamy K.R.M. (2023). "Origin, domestication, taxonomy, botanical description, genetics and cytogenetics, genetic diversity, breeding of foxtail millet (*Setaria italica* (L.) Beauv.)". *International Journal of Development Research*, 13, (09), 63660-63682.
Till-Bottraud, I., Reboud, X., Brabant, P., Lefranc, M., Rherissi, B., Vedel, F., Darmency, H. (1992). Outcrossing and hybridization in wild and cultivated foxtail millets: consequences for the release of transgenic crops. *Theoretical and Applied Genetics*, 83(8), 940-946. https://doi.org/10.1007/BF00232954
Tomar, A., Mishra, A. K., Singh, S. P., Khan, H. H., Patel, P., Singh, R. P. (2023). Nutritional benefit of small millets: Food security and sustainability in India. *Pharma Innov J*, 12, 3028-33.
Tripathi, M. K., Mohapatra, D., Jadam, R. S., Pandey, S., Singh, V., Kumar, V., Kumar, A. (2021). Nutritional composition of millets. *Millets and Millet Technology*, 101-119. https://doi.org/10.1007/978-981-16-0676-2_5
Upadhyaya, H. D., Pundir, R. P. S., Gowda, C. L. L., Reddy, V. G., Singh, S. (2009). Establishing a core collection of foxtail millet to enhance the utilization of germplasm of an underutilised crop. *Plant Genetic Resources*, 7(2), 177-184. https://doi.org/10.1017/S1479262108178042
Upadhyaya, H. D., Ravishankar, C. R., Narasimhudu, Y., Sarma, N. D. R. K., Singh, S. K., Varshney, S. K., Gowda, C. L. L. (2011). Identification of trait-specific germplasm and developing a mini core collection for efficient use of foxtail millet genetic resources in crop improvement. *Field Crops Research*, 124(3), 459-467. https://doi.org/10.1016/j.fcr.2011.08.004

Upadhyaya, H. D., Vetriventhan, M., Deshpande, S. P., Sivasubramani, S., Wallace, J. G., Buckler, E. S., Ramu, P. (2015). Population genetics and structure of a global foxtail millet germplasm collection. *The Plant Genome*, 8(3), plantgenome2015-07. https://doi.org/10.3835/plantgenome2015.07.0054

Van, K., Onoda, S., Kim, M. Y., Kim, K. D., Lee, S. H. (2008). Allelic variation of the Waxy gene in foxtail millet [*Setaria italica* (L.) P. Beauv.] by single nucleotide polymorphisms. *Molecular Genetics and Genomics*, 279(3), 255-266. https://doi.org/10.1007/s00438-007-0310-5

Vetriventhan, M., Upadhyaya, H. D., Dwivedi, S. L., Pattanashetti, S. K., Singh, S. K. (2016). Finger and foxtail millets. In *Genetic and genomic resources for grain cereals improvement* (pp. 291-319). Academic press. https://doi.org/10.1016/B978-0-12-802000-5.00007-1

Vetriventhan, M., Upadhyaya, H. D. (2019). Variability for productivity and nutritional traits in germplasm of kodo millet, an underutilised nutrient-rich climate smart crop. *Crop Science*, 59(3), 1095-1106. https://doi.org/10.2135/cropsci2018.07.0450

Vetriventhan, M., Azevedo, V. C., Upadhyaya, H. D., Nirmalakumari, A., Kane-Potaka, J., Anitha, S., Tonapi, V. A. (2020). Genetic and genomic resources, and breeding for accelerating improvement of small millets: current status and future interventions. *The Nucleus*, 63(3), 217-239. https://doi.org/10.1007/s13237-020-00322-3

Wang, C., Jia, G., Zhi, H., Niu, Z., Chai, Y., Li, W., Diao, X. (2012). Genetic diversity and population structure of Chinese foxtail millet [*Setaria italica* (L.) Beauv.] landraces. *G3: Genes| Genomes| Genetics*, 2(7), 769-777. https://doi.org/10.1534/g3.112.002907

Wang, X. Y., Diao, X. M., Wang, J. Z., Wang, C. F., Wang, G. Q., Hao, X. F., Zhao, F. F. (2013). Construction of Millet Genetic Map and QTL Analysis. *Journal of Plant Genetic Resources*, 14(5), 871-878.

Wang, Z. M., Devos, K. M., Liu, C. J., Wang, R. Q., Gale, M. D. (1998). Construction of RFLP-based maps of foxtail millet, *Setaria italica* (L.) P. Beauv. *Theoretical and Applied Genetics*, 96(1), 31-36. https://doi.org/10.1007/s001220050705

Wang, J., Wang, Z., Du, X., Yang, H., Han, F., Han, Y., Guo, E. (2017). A high-density genetic map and QTL analysis of agronomic traits in foxtail millet [*Setaria italica* (L.) P. Beauv.] using RAD-seq. *PloS One*, 12(6), e0179717. https://doi.org/10.1371/journal.pone.0179717

Wensheng, C., Yuzhen, K., Gui, D. (1991). A preliminary observation on breeding of photoperiod-sensitive and dominant GMS (Genic Male-Sterile) material of foxtail millet" Guang A1". *Acta Agriculturae Boreali-sinica* (China), 6, 47-52.

Xue, C., Zhi, H., Fang, X., Liu, X., Tang, S., Chai, Y., Diao, X. (2016). Characterization and fine mapping of SiDWARF2 (D2) in foxtail millet. *Crop Science*, 56(1), 95-103. https://doi.org/10.2135/cropsci2015.05.0331

Yadav, C. B., Muthamilarasan, M., Pandey, G., Prasad, M. (2015). Identification, characterization and expression profiling of Dicer-like, Argonaute and RNA-dependent RNA polymerase gene families in foxtail millet. *Plant Molecular Biology Reporter*, 33(1), 43-55. https://doi.org/10.1007/s11105-014-0736-y

Yoshitsu, Y., Takakusagi, M., Abe, A., Takagi, H., Uemura, A., Yaegashi, H., Terauchi, R., Takahata, Y., Hatakeyama, K., Yoko, S. (2017). QTL-seq analysis identifies two genomic regions determining the heading date of foxtail millet, *Setaria italica* (L.) P. Beauv. *Breeding Science*, 67(5), 518–527. https://doi.org/10.1270/jsbbs.17061

Yousaf, L., Hou, D., Liaqat, H., Shen, Q. (2021). Millet: A review of its nutritional and functional changes during processing. *Food Research International*, 142, 110197. https://doi.org/10.1016/j.foodres.2021.110197

Yu, J., Buckler, E. S. (2006). Genetic association mapping and genome organization of maize. *Current Opinion in Biotechnology*, 17(2), 155-160. https://doi.org/10.1016/j.copbio.2006.02.003

Zhang, G., Liu, X., Quan, Z., Cheng, S., Xu, X., Pan, S., Wang, J. (2012). Genome sequence of foxtail millet (*Setaria italica*) provides insights into grass evolution and biofuel potential. *Nature Biotechnology*, 30(6), 549-554. https://doi.org/10.1038/nbt.2195

Zhang, W., Wang, B., Liu, B., Chen, Z., Lu, G., Ge, Y., Bai, C. (2022). Trait selection for yield improvement in foxtail millet (*Setaria italica* Beauv.) under climate change in the North China plain. *Agronomy*, 12(7), 1500. https://doi.org/10.3390/agronomy12071500

Genetic Resources

ORIGINAL ARTICLE

Genetic Resources (2026), 7 (13), 181–191
DOI: 10.46265/genresj.ZUEQ4037
https://www.genresj.org
ISSN: 2708-3764

Selection of a core collection from the US castor bean germplasm collection

J. Bradley Morris[a,*], Brandon Tonnis[a], Zhenbang Chen[b] and Ming Li Wang[a,*]

[a] *USDA, ARS, Plant Genetic Resources Conservation Unit, Griffin, GA, 30223, USA*
[b] *Department of Crop and Soil Sciences, University of Georgia, Griffin, GA, 30223, USA*

Abstract: Castor bean (*Ricinus communis* L.) is a medicinal, industrial and biodiesel crop that is adapted to marginal soils in hot, dry and semi-arid environments, but its genetic potential is not fully exploited. Genetic variation exists in different castor bean genebanks, which hold many germplasm accessions. Since there are large numbers of accessions (normally over 1,000 accessions) in any genebank, the efficient way to exploit genetic diversity is to establish a core collection (i.e. 10% of the collection, but maximally representing the genetic diversity of the entire collection). There are 1,033 accessions in the United States Department of Agriculture (USDA) castor genebank, but a castor bean core collection was not available. For assessment of the genetic variation, we evaluated up to 347 accessions with available morphological and seed production data in the Germplasm Resources Information Network (GRIN) for seven qualitative and quantitative descriptors (plant height, maturity, raceme length, seed colour, seed size, stem colour and seed numbers) and then analyzed seeds chemically using nine quantitative traits (oil percentage, palmitic acid, stearic acid, oleic acid, linoleic acid, linolenic acid, gadoleic acid, ricinoleic acid and dihydrosterculic acid content).
Principal component analysis (PCA) was performed and showed that plant architecture, maturity, seed yield, oil percentage, and fatty acid profiles displayed the greatest genetic variation. Based on results from the above analysis, a core collection with 126 accessions was established. The selected accessions were classified into four groups. The results from morphological and chemical analysis were consistent. This core collection represents the genetic diversity of the entire USDA castor bean germplasm collection and can be used for genetic research and breeding improvement programmes.

Keywords: *Ricinus communis* L., core collection, principal component analysis, fatty acid, morphology

Citation: Morris, J. B., Tonnis, B., Chen, Z. and Wang, M. L. (2025) "Selection of a core collection from the US castor bean germplasm collection", *Genetic Resources*, 7(13), pp. 181–191. doi: 10.46265/genresj.ZUEQ4037

Introduction

Castor bean (*Ricinus communis* L.) is in the Euphorbiaceae family (Mubofu, 2016) and is distributed in deserts, forests, sand dunes, coastal regions, riverbeds, hill tops, valleys, roadsides, tropical, and wasteland worldwide (Anjani, 2012). Castor oil is the primary economical portion used worldwide, especially for medicinal products, as an ingredient in bath lotions and soaps, and industrial uses such as biodiesel (Morris, 2004; Senthilvel *et al*, 2017). Castor bean oil content has been shown to range from 37.2 to 60.6% (Wang *et al*, 2010) among 1,033 accessions in the USDA, Plant Genetic Resources Conservation Unit gene bank collection. An earlier study by Wang *et al*, 2011 showed that castor oil consisted of mainly eight fatty acids: 1.48% palmitic (C16:0), 1.58% stearic (C18:0), 4.41% oleic (C18:1), 6.42% linoleic (C18:2), 0.68% linolenic (C18:3), 0.45% gadoleic (C20:1), 84.51% ricinoleic (C18:1 – 1OH), and 0.47% dihydroxystearic (C18:0 – 2OH) acids (DHSA). Castor beans have important health considerations because of their use as a laxative (FDA, 2003; Morris *et al*, 2023). Castor beans have phytochemicals with potential health effects, such as ricin, which can induce cell death in cancers (Olsnes and Pihl, 1981; Calvete *et al*, 1994; Schnell *et al*, 1996; Herrera *et al*, 2003; Herrera *et al*, 2009; Park *et al*, 2022), including small-cell lung cancer (Derbyshire and Wawrzynczak, 1992; Zangemeister-Wittke *et al*, 1993). Castor oil has also been shown to be useful for improving dry eye (Khanal *et al*, 2007).

Castor bean is grown in Brazil, China, Ethiopia, India, Mozambique, Paraguay and Thailand (Singh *et al*, 2015; Landoni *et al*, 2023) with worldwide seed yields of more

* Corresponding authors: J. Bradley Morris (Brad.Morris@usda.gov), Ming Li Wang (MingLi.Wang@usda.gov)

Received: 03.02.2026 | Accepted: 20.04.2026 | Published online: 10.06.2026

than 1,3 million tonnes in 2008 (FAO, 2008) and 1.76 million tonnes in 2010 (Cheema *et al*, 2021). The crop is also cultivated in the Mediterranean area (Cafaro *et al*, 2025). Pathogens, insects, water and high temperature are the main drivers of low seed yield. Therefore, improving biotic and abiotic resistances are the main objectives for breeding programmes. Genomic technology, bioinformatics, marker trait relations and faster screening tests will improve breeding programmes (Morris *et al*, 2023).

Genetic resources are required for improving castor bean traits with agronomic importance. Since it is difficult to manage large numbers of castor bean accessions, it is crucial to develop core collections which represent the genetic variation of the entire collection. Castor bean core collections were developed in India (Sarada and Anjani, 2011; Anjani *et al*, 2018) and China (Xu *et al*, 2019). The genetic diversity and population structure of castor bean germplasm from 574 accessions of the US castor bean collection were evaluated using 22 polymorphic expressed sequence tag-simple sequence repeats (EST-SSR) markers (Wang *et al*, 2017). However, a castor bean core collection has not been developed from the United States Department of Agriculture (USDA) castor bean germplasm because the entire collection (1,033 accessions) had not been previously evaluated by EST-SSR markers. A smaller group of accessions representing the diversity of the entire germplasm collection of a species, is the method needed to strengthen the use of germplasm in cultivar development (Upadhyaya *et al*, 2014). Castor bean germplasm originating from across the globe will aid in the development of a core collection, which represents the genetic variation of the entire collection. Therefore, the main objective of this study was to establish a core collection from the entire USDA castor bean germplasm collection (1,033 accessions) based on fatty acid composition and EST-SSR and up to 347 accessions with available morphological and seed production data in GRIN using principal component analysis (PCA) and genetic diversity analysis.

Materials and methods

Plant material and morphological descriptors

Castor bean seeds used in this study originated from the entire US castor bean germplasm collection consisting of 1,033 accessions, which are maintained at the USDA, ARS, Plant Genetic Resources Conservation Unit (PGRCU), Griffin, Georgia and the National Center for Genetic Resources Preservation (NCGRP), Ft. Collins, Colorado. The regeneration field was rotovated prior to planting castor accessions. Data recorded originated from 1,033 castor accessions in cold storage at -18°C for oil content and fatty acids. However, up to 347 accessions were planted by hand or using a cone planter and grown during 1995 to 2015 in regeneration cycles and evaluated for morphological and seed production data. The additional castor accessions require evaluating in the future. Approximately 20 plants/accession were grown in each cycle. Plants were irrigated using an overhead water gun and fertilized with 20:20:20 side dressing as needed. Climatic conditions showed an average maximum and minimum temperature of 22.6°C and 10.7°C, respectively. Total precipitation was 125.88cm and 118 total rainy days.

Oil content

The seed oil content was measured for the entire castor bean collection (1,033 accessions) using a mini spec mq10 nuclear magnetic resonance (NMR) analyzer (Bruker Optics Inc., Houston, TX). All accessions were measured in triplicate to verify the precision of the NMR signal. The castor oil content was averaged from three replications. The NMR analyzer was operated at a resonance frequency of 9.95MHz and was maintained at 40°C. Our earlier results showed that even small differences in sample temperature can have a substantial effect on the NMR signal of castor oil. To avoid this effect, all samples were tempered to 40°C for 90 min before measurements.

Fatty acid profiling

The fatty acid composition of castor bean seeds was analyzed for the entire castor bean collection (1,033 accessions) by gas chromatography (GC) using a Hewlett Packard 5890 Series II gas chromatograph equipped with a flame ionization detector (FID) and an HP-7673 autosampler. A fatty acid methyl ester (FAME) standard mix RM-3 (purchased from Sigma) was used to establish peak retention times. Peak separation was performed on a DB-225 capillary column (15m × 0.25mm i.d. with a 0.25μm film) from Agilent Technologies. The carrier gas was helium set to a flow rate of 1.2ml/min. One μl of sample was injected into the column maintained isothermally at 200°C, with an injection temperature of 280°C and a detection temperature of 300°C. Total run time for each sample was 12 minutes. Fatty acid composition was determined by identifying and calculating relative peak areas. For sample preparation, 10–15 castor bean seeds were pressed by a hydraulic jack to release castor oil. One drop of castor oil was transferred into a 16 × 100mm disposable test tube, and 5.0ml of n-heptane (Fisher Scientific) was added to extract the oil. For the conversion of fatty acids to methyl esters, 500μl of 0.5 N sodium methoxide ($NaOCH_3$) in methanol solution was added to the test tube and mixed with the sample. The reaction proceeded for 2 hours, and then 7.0ml of distilled water was added to the test tube to separate the organic layer from the aqueous layer and castor bean residue (45 min). Afterwards, 1.5ml from the organic layer containing methyl esters was transferred to a 2.0ml autosampler vial for GC analysis.

SSR marker analysis

DNA was extracted from three seed tissue samples (~150mg), which had been used for fatty acid analysis. This would ensure that the generated DNA would be reliable for future oil and fatty acid association analysis. Publicly available EST-SSR markers (Qiu *et al*, 2010) were used to evaluate eight chosen castor bean accessions. Twenty-two polymorphic EST-SSR markers were used for genotyping 574 castor bean accessions.

Cluster analysis

Clustering for the core was then performed on the morphological and chemical data by entering the similarity matrix into PROC CLUSTER for cluster analysis with the unweighted paired group method using mathematical averages (UPGMA, Sneath and Sokal, 1973; Rohlf, 2000) by

specifying the AVERAGE option (SAS Institute, 2012). Following the method proposed by Nei and Li (1979), genetic diversity among 126 accessions for the core collection was calculated using NTAYA-pc software (version 2.10e, Numerical Taxonomy and Multivariate Analysis System, Rohlf, 2000). Genetic dissimilarity matrix was calculated by using SIMINT (similarity for interval data), the average genetic distance between any two accessions was calculated using the following formula:

$$DIST_{ij} = \sqrt{\sum_k \frac{1}{n}(x_{ki} - x_{kj})^2}$$

In the formula, *i* and *j* represent different accessions, *k* represents the traits investigated, and *n* represents the total number of traits investigated. Cluster analysis was conducted according to the unweighted pair group method and arithmetic average (UPGMA, Sneath and Sokal, 1973; Rohlf, 2000). Dissimilarity coefficient (Rohlf, 2000) was used to measure the genetic diversity between any two accessions.

Principal component analysis

Principal component analysis using PROC PRINCOMP (SAS Institute, 2012) was used for multivariate analysis of the data. Descriptor data has not been determined for the entire castor bean collection in the USDA, ARS, PGRCU collection, consisting of 1,033 accessions. Up to 347 accessions with descriptor characteristic data recorded from the field from 1996 to 2015 were used in the development of the core collection. Eigenvalues, the percentage of variances explained by each principal component, eigenvectors, and Pearson correlation coefficients were also determined for the core collection.

Results

Development of a core collection

The entire castor bean collection (1,033 accessions) was used in the development of a core collection by analyzing these accessions for oil content and fatty acid composition. Based on the variation for oil content (ranging from 37.2 to 60.6%) and fatty acid profiles in castor bean including 1.48% palmitic (C16:0), 1.58% stearic (C18:0), 4.41% oleic (C18:1), 6.42% linoleic (C18:2), 0.68% linolenic (C18:3), 0.45% gadoleic (C20:1), 84.51% ricinoleic (C18:1-1OH), and 0.47% dihydroxystearic (C18:0-2OH) acid profiles, the number of accessions was reduced to 574. These were genotyped with 22 SSR markers, and the number of accessions was reduced further by removing similar accessions. After the first cluster analysis using neighbour-joining analysis (Wang *et al*, 2017) with 273 accessions, 45 accessions, which showed identical branching and 5 accessions with missing data were removed from further cluster analysis. A second cluster analysis using neighbour-joining analysis (Wang *et al*, 2017) included 223 accessions, which showed a total genetic distance coefficient reduction from 798.31 to 788.27, meaning the number of accessions could be reduced by 18.3% with only a 1.2% total genetic diversity loss. After this second cluster analysis, 97 accessions were removed from the branches, including multiple accessions showing less than ten dissimilarity coefficient units. The total genetic dissimilarity coefficient among the core of 126 accessions still shows 755.45. More than 95% of the genetic diversity was retained from only 46% (126 accessions) (Figure 1) of the total 273 accessions.

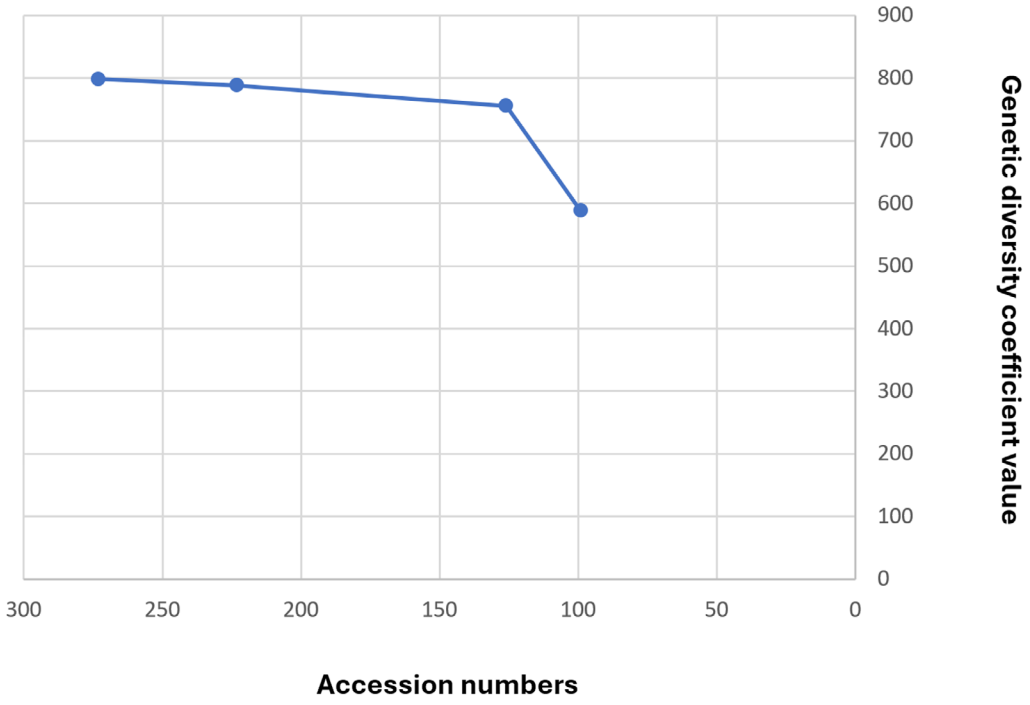


Figure 1. Size and genetic diversity coefficient value of a castor bean core collection based on oil content and fatty acid profiling at USDA, ARS, PGRCU

A further reduction of 27 accessions would significantly reduce the genetic diversity by 23%. Therefore, these 126 accessions will be utilized for core development in the castor bean collection. This castor bean core collection (Supplemental Table 1) represents accessions from Afghanistan (1), Algeria (3), Argentina (4), Benin (1), Brazil (9), Bulgaria (1), Colombia (1), Cuba (2), Democratic Republic of the Congo (1), Egypt (2), Greece (1), India (36), Iran (16), Jordan (1), Kenya (1), Mexico (2), Morocco (2), Pakistan (1), Panama (1), Paraguay (1), Peru (2), Serbia (1), South Africa (14), former Soviet Union (3), Syria (1), Turkey (13), United States (3), and Uruguay (2).

The scheme showing how the castor core collection was selected is outlined below:

A. Chemical analysis based on oil content and fatty acid profiles (1,033 accessions).
B. Based on the variation for oil content and fatty acid profiles, the number of accessions was reduced to 574 and these were genotyped with 22 SSR markers.
C. First cluster analysis (273 accessions).
D. 50 accessions showing identical branching and missing data were removed.
E. Second cluster analysis (223 accessions).
F. 97 accessions removed from branches including multiple accessions with <10 dissimilarity coefficient units.
G. 95% of genetic diversity retained (126 accessions used to form the core collection).
H. Confirmation using principal component and cluster analysis of chemical, morphological, and reproductive traits (up to 347 accessions) was used to verify the core collection of 126 accessions.

Descriptor data

Castor bean descriptor data in the Germplasm Resources Information Network (GRIN) (USDA, 2024) was used for recording qualitative and quantitative observations on up to 347 accessions over the period of 1995 - 2015. Plant characteristics for plant height (dm), maturity (early, midseason, late), raceme length (cm), and stem colour (green, red, mixture of green and red) were recorded from an average of 20 plants/accession at 50% maturity. Seed colour (brown, tan, reddish brown), seed size (small, medium, large) and seed numbers were recorded from all plants after seed drying. The qualitative characteristics were measured according to the scales shown in Supplemental Table 2. The data may represent different accessions for each of the descriptors and GRIN may not be up to date.

Confirmation of diversity of the core collection

Confirmation analysis using principal component and cluster analysis was used to verify this core collection. The number of accessions used in this analysis consisted of 262 to 347 accessions for the original collection. This range was necessary because not all of the original castor accessions had been characterized for morphological, reproductive and phenological data. The means for morphological, phenological and reproductive traits in the core and original accessions were similar (Table 1).

Table 1. Comparison of means, ranges, and variances for seven morphological, phenological, and reproductive traits in the core and original accessions after removal of similar accessions based on EST-SSR markers. SE, Standard error; SD Standard deviation.

		Original collection				
Trait	**N**	**Mean ± SE**	**SD**	**Range**	**Variance**	**T-value**
Plant height (cm)	285	21 ± 0.54	9.05	5–43	81.9	< 0.0001
Maturity	293	4.2 ± 0.19	3.19	1–9	10.2	< 0.0001
Raceme length (cm)	262	27.6 ± 0.71	11.49	10–61	132.1	< 0.0001
Seed colour	275	1.6 ± 0.05	0.83	1–3	0.69	< 0.0001
Seed size	273	4.2 ± 0.14	2.28	1–9	5.2	< 0.0001
Stem colour	295	1.6 ± 0.04	0.64	1–3	0.42	< 0.0001
Seed number	347	523 ± 30.1	560.26	44–4321	313,898	< 0.0001
		Core collection				
Trait	**N**	**Mean ± SE**	**SD**	**Range**	**Variance**	**T-value**
Plant height (cm)	126	22 ± 0.85	9.5	5–43	90.3	<0.0001
Maturity	126	4.1 ± 0.27	3.08	1–9	9.5	<0.0001
Raceme length (cm)	126	29.3 ± 1.11	12.47	10–61	155.5	<0.0001
Seed colour	126	2 ± 0.08	0.84	1–3	0.72	<0.0001
Seed size	126	4 ± 0.21	2.35	1–9	5.5	<0.0001
Stem colour	126	2 ± 0.06	0.62	1–3	0.38	<0.0001
Seed number	126	750 ± 69.2	776.83	44–4321	603469	<0.0001

The range of mean values were identical in the core when compared to the original accessions. Variances between the core and original accessions (262 to 347 accessions), except for seed number, were homogenous, indicating that the diversity of the original accessions was represented in the core collection (Table 1). The correlation patterns for morphological, phenological and reproduction traits (Table 2) were similar in the original accessions and core collection, showing that associations observed in the original accessions were well represented in the core collection.

Six of the 21 correlation coefficients in the core collection were significant ($P = 0.01$, $P = 0.05$, $P = 0.001$) and ranged from 0.2 to 0.37.

The means and ranges for oil percentage and fatty acid traits were nearly identical between the core and original accessions (Table 3).

The variances for oil percentage and fatty acid traits were homogenous, indicating that diversity from the original accessions was captured in the core collection (Table 3). The correlation pattern for oil percentage and fatty acid traits (Table 4) was similar in the original accessions and core collection, showing that associations observed in the original accessions were well represented in the core collection.

Table 2. Pearson correlation coefficients for morphological, phenological, and reproductive traits in the core and original castor bean core collection. Above the diagonal are correlation coefficients involving core collection accessions; below the diagonal are correlation coefficients involving the original accessions after removal of similar phytochemical accessions based on EST-SSR markers. *, Significant at $p = 0.05$; **, Significant at $p = 0.01$; ***, Significant at $p = 0.001$.

Trait	Plant height	Maturity	Raceme length	Seed colour	Seed size	Stem colour	Seed number
Plant height		-0.01	0.25**	-0.1	-0.0008	0.04	-0.08
Maturity	-0.1		0.12	0.11	0.02	-0.04	0.02
Raceme length	0.29***	0.15*		-0.01	0.01	0.04	-0.05
Seed colour	-0.04	0.13*	0.04		0.37***	0.2*	-0.21*
Seed size	0.03	-0.02	0.04	0.29***		0.21*	-0.22**
Stem colour	0.05	-0.06	0.06	0.19**	0.24***		-0.02
Seed number	-0.01	-0.01	-0.01	-0.14*	-0.2**	0.02	

Table 3. Comparison of means, ranges, and variances for nine oil and fatty acid traits in the core and original accessions after removal of similar accessions based on EST-SSR markers.

		Original collection				
Trait	**N**	**Mean ± SE**	**SD**	**Range**	**Variance**	**T-value**
Oil %	347	48.7 ± 0.14	2.72	35–61	7.41	> 0.0001
Palmitic (16:0)	346	1.5 ± 0.01	0.19	1.04–2.44	0.03	> 0.0001
Stearic (18:0)	346	1.6 ± 0.02	0.42	0.87–4.6	0.18	> 0.0001
Oleic (18:1)	346	4.9 ± 0.06	1.12	2.41–9.77	1.25	> 0.0001
Linoleic (18:2)	346	6.4 ± 0.03	0.61	4.76–8.79	0.38	> 0.0001
Linolenic (18:3)	346	0.7 ± 0.009	0.16	0.46–1.45	0.02	> 0.0001
Gadoleic (20:1)	346	0.5 ± 0.007	0.13	0.18–0.91	0.01	> 0.0001
Ricinoleic	346	83.9 ± 0.08	1.58	78.59–87.73	2.52	> 0.0001
Dihydroxystearic	346	0.5 ± 0.005	0.1	0.19–0.92	0.01	> 0.0001
		Core collection				
Trait	**N**	**Mean ± SE**	**SD**	**Range**	**Variance**	**T-value**
Oil %	126	48.9 ± 0.28	3.19	35–61	10.18	< 0.0001
Palmitic (16:0)	126	1.5 ± 0.2	0.17	1.14–2.12	0.03	< 0.0001
Stearic (18:0)	126	1.6 ± 0.03	0.33	0.87–3	0.11	< 0.0001
Oleic (18:1)	126	5 ± 0.09	1	3.05–8.26	1	< 0.0001
Linoleic (18:2)	126	6.4 ± 0.05	0.6	4.76–8.68	0.36	< 0.0001
Linolenic (18:3)	126	0.7 ± 0.01	0.15	0.46–1.41	0.02	< 0.0001
Gadoleic (20:1)	126	0.5 ± 0.01	0.13	0.22–0.9	0.01	< 0.0001
Ricinoleic	126	83.8 ± 0.12	1.35	79.05–86.99	1.84	< 0.0001
Dihydroxystearic	126	0.5 ± 0.008	0.09	0.27–0.74	0.008	< 0.0001

Table 4. Pearson correlation coefficients for oil percentage and fatty acid traits in the core and original accessions after removal of similar phytochemical accessions based on EST-SSR markers. Above the diagonal are correlation coefficients involving core collection accessions; below the diagonal are correlation coefficients involving the original accessions after removal of similar phytochemical accessions based on EST-SSR. *, Significant at $p = 0.05$; **, Significant at $p = 0.01$; ***, Significant at $p = 0.0001$.

Trait	Oil %	16:0	18:0	18:1	18:2	18:3	20:1	Ricinoleic	DHSA
Oil %		-0.04	-0.05	0.17*	0.11	-0.31**	0.05	-0.13	0.004
16:0	0.02		0.18*	-0.04	0.7***	0.25**	-0.05	-0.42***	-0.53***
18:0	-0.03	0.23***		0.29**	-0.03	-0.11	0.05	-0.43***	-0.31**
18:1	0.1*	-0.02	0.36***		-0.01	-0.42***	0.45***	-0.76***	-0.33**
18:2	0.06	0.71***	0.04	0.01		0.1	0.32**	-0.53***	-0.12
18:3	-0.28***	0.32***	0.09	-0.39***	0.16**		-0.16	0.18*	-0.06
20:1	-0.006	-0.04	0.12*	0.51***	0.32***	-0.11*		-0.56***	0.1
Ricinoleic	-0.06	-0.44***	-0.57***	-0.79***	-0.53***	0.06	-0.59***		0.37***
DHSA	-0.04	-0.54***	-0.26***	-0.25***	-0.15**	-0.18**	0.13*	0.32***	

Eighteen of the 36 correlation coefficients in the core collection were significant ($P = 0.01$, $P = 0.05$, $P = 0.0001$) and were in the range from 0.17 to 0.76.

Cluster analysis

Average distance cluster analysis from morphological data grouped the core castor accessions into four well-defined clusters with distinct plant height, raceme length and seed numbers (Supplemental Figure 1).

Twelve accessions in clusters 1 and 2 showed the shortest plants (mean of 16dm) and raceme length (mean of 21cm), and the highest mean seed numbers (2,949). Thirty-nine accessions in cluster 3 showed the tallest plants (25dm), the second-longest raceme length (29cm), and medium seed numbers (1,021). Seventy-five accessions in cluster 4 showed a mean of medium height (21dm), longest raceme length (30cm), and the lowest seed number (305).

Average distance cluster analysis from the chemical data grouped the core castor accessions into four well-defined clusters with distinct oil content and fatty acid concentrations. (Supplemental Figure 2).

Cluster 1 had three accessions and showed the highest DHSA content (0.52%), second-highest stearic, oleic, gadoleic acid content (1.8, 5.1, 0.52%, respectively), and the lowest linoleic acid and linolenic acid content (6.1 and 0.61%, respectively). Clusters 1 and 3 showed the highest ricinoleic acid content (84%), and the lowest oil content (45%). Cluster 2 (2 accessions) showed the highest oil, stearic, oleic, linoleic, gadoleic acid content (49, 2.2, 8.2, 6.8, and 0.59%, respectively), and the lowest mean DHSA and ricinoleic acid content (0.36 and 79.6% respectively). Cluster 3 (29 accessions) showed the highest linolenic acid content (0.75%), the third highest mean stearic, linoleic, DHSA content (1.7, 6.4, 0.47%, respectively), and lowest gadoleic acid content (0.46%). Cluster 4 (92 accessions) showed the second highest mean oil, linoleic, and DHSA content (48, 6.5 and 0.5%, respectively), lowest mean stearic acid (1.6%), third highest gadoleic, and DHSA content (0.64 and 0.5, respectively). Clusters 3 and 4 had the third-highest oleic acid content (5%). Clusters 2 and 4 had the second-highest linolenic acid content (0.64%). These results show variation for oil, stearic, oleic, linoleic, linolenic, gadoleic, ricinoleic acid, and DHSA seed traits in castor beans. Average distance cluster analysis was also performed using both morphological and chemical data (Figure 2). The core collection was still grouped into four clusters with the same accession numbers for each group (Cluster 1 with 4 accessions, Cluster 2 with 8 accessions, Cluster 3 with 39 accessions, and Cluster 4 with 75 accessions). However, some accession listed orders were switched within each cluster. This means that our cluster analysis was very robust and consistent.

Principal component analysis

Phenotypic, maturity, and seed reproduction principal component analysis accounted for 24% and 22% of the total variation at the first principal component in the core and the original accessions, respectively (Table 5).

The amount of variation accounted for, cumulatively, by adding principal components 2 through 6 was 43, 58, 72, 82, 91%, and 41, 55, 68, 78, 87% for the core and original castor bean accessions, respectively. The first principal component was most correlated with seed colour and seed size in the core and plant height and raceme length in the original accessions (Table 6).

The second principal component accounted for 19% of the variation in the core and original accessions, and was mostly due to plant height and raceme length in the core, while seed colour and seed size were mostly correlated in the original accessions. The third principal component explained 15% and 14% of the variation in the core and original accessions, respectively, and was composed primarily of maturity. The fourth principal component accounted for 14% and 13% of the variation in the core and original castor accessions and was most correlated with stem colour and seed number. The fifth principal component explained 10% of the variation in the core and original accessions and mostly correlated with plant height and raceme length in the core and seed size in the original accessions. The sixth principal component explained 10% and 9% of the variation in the core and original accessions and was composed primarily of seed size, stem colour, and seed number in the core. However, seed colour and stem colour were mostly correlated in the original accessions. Therefore, castor bean cultivars from this core collection could be developed with improved architecture, early or late maturity, and low or high seed yield.

Percent oil and fatty acids accounted for 33% and 34% of the total variation in the core and original accessions, respectively, at the first principal component (Table 5). The cumulative amount of variation for components 2 through 6

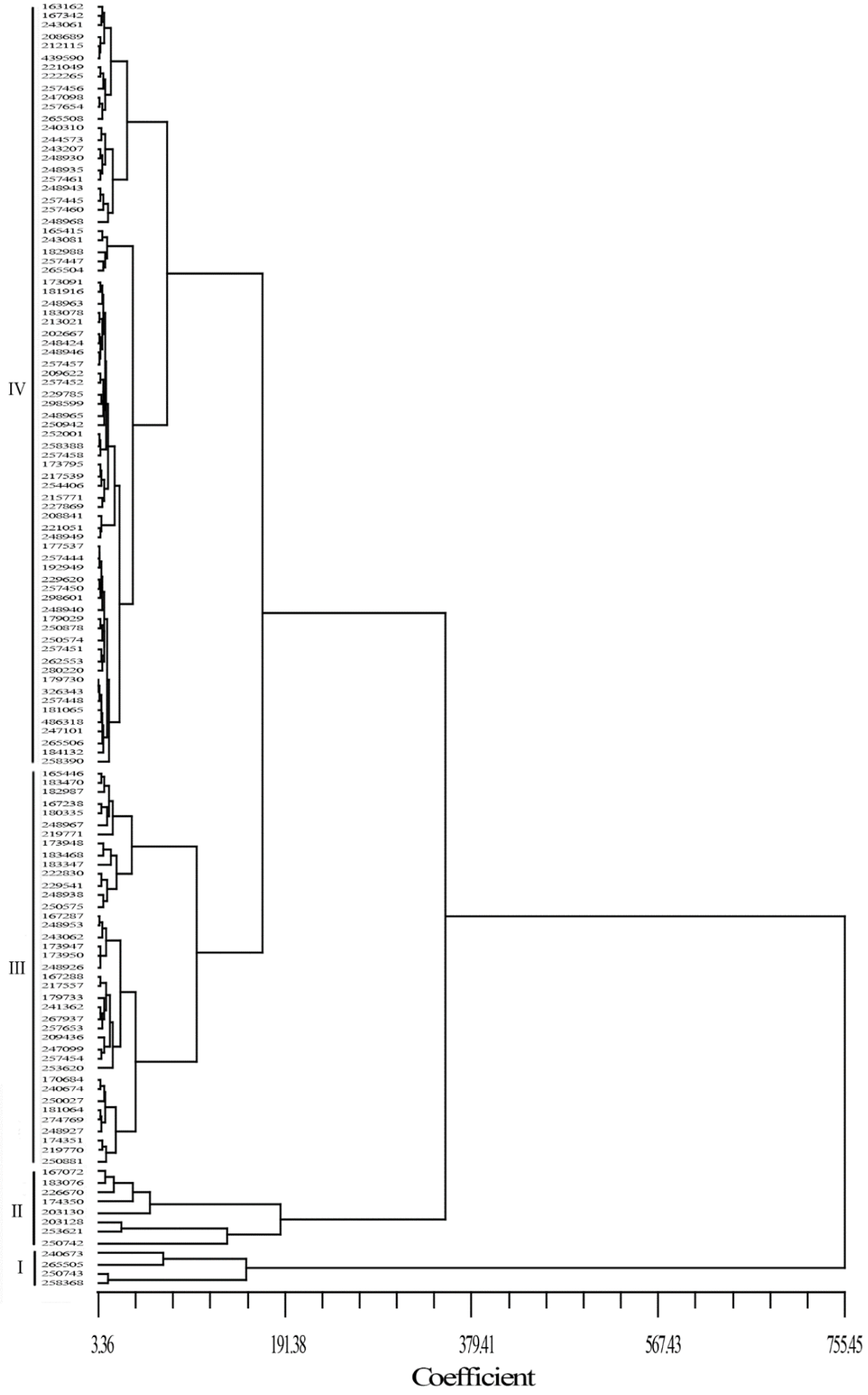


Figure 2. Unweighted pair group method using mathematical averages (UPGMA) dendrogram of 126 selected core accessions for chemical and morphological data.

Table 5. Eigenvalues and the proportion of total phenotypic, maturity, seed reproduction, oil, and fatty acid variability among the core and original castor bean accessions as explained by the principal components.

Principal component	Eigenvalue		% Variability		% Cumulative	
	Core	Original	Core	Original	Core	Original
Phenotypic, maturity, and seed reproduction						
1	1.677	1.749	23.96	21.86	23.96	21.86
2	1.3014	1.5019	18.59	18.77	42.55	40.64
3	1.0744	1.1584	15.35	14.48	57.9	55.12
4	0.9739	1.0328	13.91	12.91	71.81	68.03
5	0.7032	0.8104	10.05	10.13	81.86	78.16
6	0.6737	0.6983	9.63	8.73	91.49	86.89
Oil and fatty acids						
1	2.9463	3.0218	32.74	33.58	32.74	33.58
2	1.9622	2.0021	21.8	22.25	54.54	55.82
3	1.348	1.2053	14.98	13.39	69.52	69.21
4	1.0663	1.1695	11.85	12.99	81.37	82.21
5	0.6688	0.6791	7.43	7.55	88.8	89.76
6	0.57	0.5461	6.33	6.07	95.13	95.82

was 55% to 96% in the core and original accessions. The first principal component was mostly correlated with ricinoleic acid in the core and original accessions, while the second principal component correlated with palmitic and linolenic acid (Table 6). The third principal component was mostly correlated with stearic and linoleic acid. The fourth principal component was primarily correlated with oil percentage, while the fifth principal component correlated mostly with stearic acid. The sixth principal component mostly correlated with linolenic acid. Therefore, potential exists to develop castor bean cultivars from this core collection with improved oil percentage and fatty acid profiles.

Discussion

Genebanks worldwide curate collections containing many accessions and may restrict uses of genetic diversity (Gireesh *et al*, 2023). Genetic variation in genetic resource collections is essential for plant breeding. Since large germplasm collections may inhibit use, smaller core collections could be developed from these larger collections to represent diversity for target traits (Frankel, 1984). Core collections were developed in sugarcane (*Saccharum spontaneum* L.) (Tai and Miller, 2001), *Medicago* (Diwan *et al*, 1994), barley (*Hordeum vulgare* L.) (Yuan *et al*, 2024), sorghum (*Sorghum bicolor* L.) (Upadhyaya *et al*, 2009), peanut (*Arachis hypogaea* L.) (Holbrook and Dong, 2005), and foxtail millet (*Setaria italica* L.) (Choi *et al*, 2018). A barley core collection was shown to consist of disease-resistant loci and can be used in strengthening barley hardiness (Yuan *et al*, 2024). Comparing means, variances, and correlation coefficients for the castor bean traits in the entire and core collections (Tables 1, 2, 3 and 4) in this study show that genetic variation has been preserved in the core collection containing 126 accessions. This core collection, developed based on fatty acid profiling, provides sources of variation useful in castor bean improvement. Multivariate analysis, including cluster and principal component analysis, is useful for constructing core collections (van Hintum *et al*, 2000; Ruperao, 2024), and have been used on sugarcane (Tai and Miller, 2001), perennial *Medicago* (Basigalup *et al*, 1995), sesame (*Sesamum indicum* L.) (Yol and Uzun, 2012) and tea (*Camelliasinensis* L. O. Kuntze) (Kottawa-Arachchi *et al*, 2024).

Generally, core collections should be 10% of the whole collection and less than 2,000 accessions and contain approximately 70% of the variation (Brown, 1989). However, some core collections represent 5–10% showing 75–90% of the variation (Charmet and Balfourier, 1995; Bisht *et al*, 1998). Core collections have ranged in size from 600 (Gireesh *et al*, 2023) to 2,500 (van Hintum *et al*, 2000) accessions representing 1.5% and 31% for sorghum and *Solanum,* respectively. This castor core collection containing 126 accessions, represents 12% of the entire collection, which is an ideal number.

Conclusion

The USDA castor bean germplasm collection was characterized by morphological descriptors for up to 347 accessions with available data in GRIN and analyzed chemically for fatty acid composition from 1,033 accessions. These accessions were clustered into four groups from morphological and chemical data. Part of the collection (574 accessions) was genotyped using EST-SSR markers and clustered into four groups. The morphological, chemical and genetic data supported each other. Based on the results from the above analysis, 126 accessions were selected to form the core collection (12% of the entire collection), which well represented the genetic diversity of the entire collection. The accessions in the core collection are freely available and can be requested online for genetic research and breeding programmes by the castor research community.

Table 6. Eigenvectors, principal components for seven phenotypic, maturity, seed, and nine oil, fatty acid traits in the core and original castor bean accessions.

Original collection						
	Principal components					
Trait	1	2	3	4	5	6
Plant height (dm)	0.57	-0.007	-0.21	-0.23	0.08	0.36
Maturity	-0.12	0.23	0.76	0.2	0.1	-0.05
Raceme length (cm)	0.52	0.05	0.28	0.09	0.4	-0.33
Seed colour	-0.11	0.61	0.05	0.15	-0.27	0.52
Seed size	-0.04	0.55	-0.25	-0.14	0.65	0.05
Stem colour	0.17	0.36	-0.39	0.52	-0.28	-0.5
Seed number	0.02	-0.34	-0.09	0.75	0.33	0.42
Oil %	0.05	-0.15	0.1	-0.77	0.48	0.35
Palmitic (16:0)	0.35	0.48	0.09	-0.19	-0.08	-0.23
Stearic (18:0)	0.33	-0.01	-0.48	0.19	0.62	-0.31
Oleic (18:1)	0.38	-0.43	-0.2	0.004	-0.25	0.13
Linoleic (18:2)	0.35	0.28	0.55	-0.08	-0.006	-0.23
Linolenic (18:3)	0.008	0.51	-0.03	0.37	0.3	0.61
Gadoleic (20:1)	0.31	-0.32	0.41	0.31	0.008	0.39
Ricinoleic	-0.55	0.12	-0.01	-0.08	-0.03	0.03
Dihydrosterculic	-0.28	-0.28	0.46	0.26	0.45	-0.32
Core collection						
	Principal components					
Trait	1	2	3	4	5	6
Plant height (dm)	0.02	0.67	-0.28	-0.03	0.63	0.11
Maturity	0.09	0.16	0.86	0.1	0.32	-0.25
Raceme length (cm)	0.09	0.67	0.15	0.11	-0.68	0.14
Seed colour	0.57	-0.18	0.19	0.005	-0.02	0.1
Seed size	0.58	-0.07	-0.04	-0.007	0.11	0.59
Stem colour	0.37	0.01	-0.31	0.69	0.006	-0.5
Seed number	-0.4	-0.12	0.14	0.69	0.09	0.52
Oil %	0.11	-0.23	0.24	-0.75	0.07	0.54
Palmitic (16:0)	0.31	0.54	0.03	-0.17	0.07	-0.2
Stearic (18:0)	0.27	-0.02	-0.51	0.13	0.73	0.24
Oleic (18:1)	0.41	-0.37	-0.17	0.06	-0.37	0.02
Linoleic (18:2)	0.33	0.35	0.5	-0.06	0.2	-0.17
Linolenic (18:3)	-0.13	0.49	0.01	0.31	-0.21	0.73
Gadoleic (20:1)	0.31	-0.25	0.4	0.45	-0.06	0.17
Ricinoleic	-0.55	0.05	-0.04	-0.12	0.009	-0.06
Dihydrosterculic	-0.31	-0.25	0.46	0.23	0.47	0.01

Supplemental data

Supplemental Table 1. Relevant morphological, seed production, oil content, and fatty acid data from 126 castor bean accessions selected for the core collection.

Supplemental Table 2. Descriptive scales of qualitative traits in the entire collection.

Supplemental Figure 1. Unweighted pair group method using mathematical averages (UPGMA) dendrogram of 126 selected core accessions for morphological and seed number data.

Supplemental Figure 2. Unweighted pair group method using mathematical averages (UPGMA) dendrogram of 126 selected core accessions for % oil and fatty acid data.

Author contributions

J. Bradley Morris: Conceptualization; analysis; methodology; writing. **Brandon Tonnis:** Methodology. **Zhenbang Chen:** Analysis; methodology. **Ming Li Wang:** Conceptualization; analysis; methodology; writing.

Acknowledgments

Authors acknowledge support from Lee Ann Chalkley for providing castor bean seeds; and Ken Manley for planting, maintenance, and harvesting field plots. USDA is an equal opportunity provider and employer.

Conflict of interest statement

The authors have declared that no competing interests exist.

References

Anjani, K. (2012). Castor genetic resources: a primary gene pool for exploitation. *Ind. Crops &Prod.* 35, 1-14. https://doi.org/10.1016/j.indcrop.2011.06.011.

Anjani, K., Raoof, M.A., Prasad, M.S.L., Duraimurugan, P., Lucose, C., Yadav, P., Prasad, R.D., Lal, J.J., Sarada, C. (2018). Trait-specific accessions in global castor (*Ricinus communis* L.) germplasm core set for utilization in castor improvement. *Ind.l Crops & Prod.* 112, 766-774. https://doi.org/10.1016/j.indcrop.2018.01.002.

Basigalup, D.H., Barnes, D.K., Stucker, R.E. (1995). Development of a core collection for perennial *Medicago* plant introductions. *Crop Sci.* 35, 949-1244. https://doi.org/10.2135/cropsci1995.0011183X003500040042xop.

Bisht, I.S., Mahajan, R.K., Lokknathan, T.R., Agrawal, R.C. (1998). Diversity in Indian sesame collection and stratification of germplasm accessions in different diversity groups. *Genet. Resour. and Crop Evol.* 45, 325-335. http://dx.doi.org/10.1023/A:1008652420477

Brown, A.H.D. (1989). The case for core collections. In *The use of plant genetic resources*, eds. A.H.D. Brown, O.H. Frankel, D.R. Marshall, & J.T. Williams (Cambridge University Press), 136-156.

Cafaro, V., Testa, G., Patanè, C. (2025). Castor: A renewed oil crop for the Mediterranean environment. Agron. 15, 1402. https://doi.org/10.3390/agronomy15061402.

Calvete, J.A., Newell, D.R., Wright, A.F., Rose, M.S. (1994). In vitro and in vivo antitumor activity of ZENECA ZD0490, a recombinant ricin A-chain immunotoxin for the treatment of colorectal cancer. *Cancer Res*. 54, 4684-4690. PMID: 8062265.

Charmet, G., Balfourier, F. (1995). The use of geostatistics for sampling a core collection of perennial ryegrass populations. *Genet. Resour. and Crop Evol*. 42, 303-309. https://doi.org/10.1007/BF02432134.

Cheema, N.M., Shabbir, G., Nawaz, N. (2021). Quality characteristics of castor bean oil (*Ricinus communis* L.) under sub-tropical divergent pothwar environment of Pakistan. *Pakistan J. Agric. Res.* 34, 813-818. https://dx.doi.org/10.17582/journal.pjar/2021/34.4.813.818.

Choi, Y.-M., Kim, K.-M., Lee, S., Oh, S., Lee, M-C. (2018). Development of a core collection based on EST-SSR markers and phenotypic traits in foxtail millet [*Setaria italica* (L.) P. Beauv.]. *J. Crop Sci. and Biotech.* 21, 395-405. https://doi.org/10.1007/s12892-018-0189-0

Derbyshire, E.J., Wawrzynczak, E.J. (1992). An anti-mucin immunotoxin BrE-3-ricin A-chain is potently and selectively toxic to human small-cell lung cancer. *Internat. J. Cancer* 52, 624-630. https://doi.org/10.1002/ijc.2910520422

Diwan, N., Bauchan, G.R., McIntosh, M.S. (1994). A core collection for the United States annual *Medicago* germplasm collection. *Crop Sci.* 34, 279-285. https://doi.org/10.2135/cropsci1994.0011183X003400010051x

FDA. (2003). *OTC Drug review ingredient report.* Food and Drug Administration, Washington, DC.

Food and Agriculture Organization of the United Nations (2008). *World crop production statistics*. FAOSTAT statistical database, Rome.

Frankel, O.H. (1984). Genetic perspectives of germplasm conservation. In *Genetic manipulation: Impact on man and society*, eds. W.K. Arber, K. Llimensee, W.J. Peacock, and P. Starlinger (Cambridge University Press), 161-170.

Gireesh, C., Anantha, M.S., Senguttuvel, P., Basavaraj, K., Rathod, S., Raveendra, C., Uday, V., Subba Rao, L.V., Sundaram, R.M. (2023). Development of core collection for plant genetic resource management and utilization. In *Advanced statistical tools and techniques for biometrical data analysis,* eds. S. Rathod, B. Sailaja, N. Bandumula, S.A. Kumar, P. Jeyakumar, A. Waris, P. Muthuraman, & R.M. Sundaram (ICAR – Indian Institute of Rice Research, Hyderabad), 197-206.

Herrera, L., Yarbrough, S., Ghetie, V., Aquino, D.B., Vitetta, E.S. (2003). Treatment of SCID/Human B cell precursor ALL with anti-CD19 and anti-CD22 immunotoxins. *Leukemia* 17, 334-338. https://doi.org/10.1038/sj.leu.2402790.

Herrera, L., Bostrom, B., Gore, L., Sandler, E., Lew, G., Schlegel, P.G., Aquino, V., Ghetie, V., Vitetta, E.S., Schindler, J. (2009). A phase 1 study of combotox in pediatric patients with Refractory B-lineage acute lymphoblastic leukemia. *J. Ped. Hematol./Oncol.* 31, 936-941. DOI: https://doi.org/10.1097/MPH.0b013e3181bdf211.

Holbrook, C.C., Dong, W. (2005). Development and evaluation of a mini core collection for the U.S. peanut germplasm collection. *Crop Sci.* 45, 1203-1684.https://doi.org/10.2135/cropsci2004.0368.

Khanal, S., Tomlinson, A., Pearce, E.I., Simmons, P.A. (2007). Effect of an oil-in-water emulsion on the tear physiology of patients with mild to moderate dry eye. *Cornea* 26, 175-181. https://doi/10.1097/ICO.0b013e31802b492d.

Kottawa-Arachchi, J., Ranatunga, M.A.B., Sharma, R.K., Chaudhary, H.K., Attanayake, R.N., Amarakoon, A.M.T., Gunasekare, M.T.K., Sharma, B., Kumar, N., Sood, V.K. (2024). Morpho-molecular genetic diversity and population structure analysis to enrich core collections in tea [*Camellia sinensis* (L.) Kuntze] germplasm of Sri Lanka and India. *Genet. Resour. Crop Evol.* 71, 2597-2616. https://doi.org/10.1007/s10722-023-01792-5

Landoni, M., Bertagnon, G., Ghidoli, M., Cassani, E., Adani, F., Pilu, R. (2023). Opportunities and challenges of castor bean (*Ricinus communis* L.) genetic improvement. *Agron.* 13, 8. https://doi.org/10.3390/agronomy13082076.

Morris, J.B. (2004). Phytochemical traits in the genetic resources of castorbean. *Current Topics in Plant Biol.* 5, 63-67.

Morris, J.B., Dierig, D., Heinitz, C., Hellier, B., Bradley, V., Marek, L. (2023). Vulnerability of U.S. new and industrial crop genetic resources. *Industrial Crops & Products* 206, 117364. https://doi.org/10.1016/j.indcrop.2023.117364.

Mubofu, E.B. (2016). Castor oil as a potential renewable resource for the production of functional materials. *Sustainable Chemical Processes* 4, 1-12. Doi https://doi.org/10.1186/s40508-016-0055-8

Nei, M., Li, W. H. (1979). Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proceedings of the National Academy of Science USA* 76, 5269–5273. https://doi.org/10.1073/pnas.76.10.5269.

Olsnes, S., Pihl, A. (1981). Chimeric toxins. *Pharmacol. & Therapeut.* 15, 355-381. https://doi.org/10.1016/0163-7258(81)90050-4.

Park, S.G., Kim, H., Jun, H., Choi, S.Y., Kim, E., Kang, S. (2022). Directing ricin-based immunotoxins with targeting affibodies and KDEL signal peptide to cancer cells effectively induces apoptosis and tumor suppression. *J. of Nanobiotech.* 20, 1-17. https://doi.org/10.1186/s12951-022-01601-8.

Qiu, L., Yang, C., Tian, B., Yang, J.B., Liu, A. (2010). Expoloiting EST databases for the development and characterization of EST-SSR markers in castor bean (*Ricinus communis* L.). *BMC Plant Biol.* 10, 278. http://www.biomedcentral.com/1471-2229/10/278.

Rohlf, J. (2000). Numerical taxonomy and multivariate analysis system Version 2.10e. Department of Ecology and Evolution, Sata University of New York at Stony Brook. Manual Applied Biostatistics, Inc N.Y.

Ruperao, P. (2024). Development of a core set from large germplasm collections in genebank. In *Bioinformatics for plant research and crop breeding,* ed. J.T. Chen (John Wiley & Sons Ltd.), 269-282. https://doi.org/10.1002/9781394209965.ch11.

Sarada, C., Anjani, K. (2011) Self-organizing mapping (SOM) networks for the development of castor core collection. In: Abstracts annual conference, society of statistics computers applications. Society of Statistics Computers Applications, 23 p.

SAS Institute Inc. (2012). SAS/STAT 9.3 User's Guide. Cary, NC, USA: SAS Institute Inc.

Schnell, R., Katouzi, A.A., Linnartz, C., Schoen, G., Drillich, S., Hansmann, M.-L., Schiefer, D., Barth, S., Zangemeister-Wittke, U., Stahel, R.A., Diehl, V., Engert, A. (1996). Potent anti-tumor effects of an anti-CD24 ricin A-chain immunotoxin *in vitro* and in a disseminated human Burkitt's lymphoma model in SCID mice. *Internat. J. Cancer* 66, 526-531. https://doi.org/10.1002/(sici)1097-0215(19960516)66:4%3C526::aid-ijc17%3E3.0.co;2-5

Senthilvel, S., Shaik, M., Anjani, K., Shaw, R.K., Kumari, P., Sarada, C., Kiran, B.U. (2017). Genetic variability and population structure in a collection of inbred lines derived from a core germplasm of castor. J. Plant Biochem. Biotechnol. 26, 27-34. https://doi.org/10.1007/s13562-016-0356-8

Singh, A.S., Kumari, S., Modi, A.R., Gajera, B.B., Narayanan, S., Kumar, N. (2015). Role of conventional and biotechnological approaches in genetic improvement of castor (*Ricinus communis* L.) *Ind. Crops & Prod.* 74, 55-62. https://doi.org/10.1016/j.indcrop.2015.05.001

Sneath, P. H., Sokal, R. R. (1973). Numerical taxonomy: The principles and practice of numerical classification. (San Francisco: Freeman), 573 p.

Tai, P.Y.P., Miller, J.D, (2001). A core collection for *Saccharum spontaneum* L. from the world collection of sugarcane. *Crop Sci.* 41, 879-885. https://doi.org/10.2135/cropsci2001.413879x.

Upadhyaya, H.D., Pundir, R.P.S., Dwivedi, S.L., Gowda, L.L., Reddy, V.G., Singh, S. (2009). Developing a mini core collection of sorghum for diversified utilization of germplasm. Crop Sci. 49, 1539-1939. https://doi.org/10.2135/cropsci2009.01.0014.

Upadhyaya, H.D., Dwivedi, S.L., Singh, S.K., Singh, M., Vetriventhan, M., Sharma, S. (2014). Forming core collections in barnyard, kodo, and little millets using morphoagronomic descriptors. *Crop Sci.* 54, 2673-2682. https://doi.org/10.2135/cropsci2014.03.0221.

USDA, Agricultural Research Service, National Plant Germplasm System. Data from: Germplasm Resources Information Network (GRIN-Global). National Germplasm Resources. (2024) https://npgsweb.ars-grin.gov/gringlobal/search.aspx.

van Hintum, Th.J.L., Brown, A.H.D., Spillane, C., Hodgkin, T. (2000). Core collections of plant genetic resources (Rome, Italy: IPGRI Technical Bulletin No. 3), 51 p.

Wang, M.L., Morris, J.B., Pinnow, D.L., Davis, J., Raymer, P., Pederson, G.A. (2010). A survey of the castor oil content, seed weight and seed-coat colour on the United States Department of Agriculture germplasm collection. *Plant Gen. Res.:Charact. and Util.* 8, 229-231.https://doi.org/10.1017/S1479262110000262.

Wang, M.L., Morris, J.B., Tonnis, B., Pinnow, D., Davis, J., Raymer, P., Pederson, G.A. (2011). Screening of the entire USDA castor germplasm collection for oil content and fatty acid composition for optimum biodiesel production. *J. of Agric. and Food Chem.* 59, 9250-9256. https://doi.org/10.1021/jf202949v.

Wang, M. L., Dzievit, M., Chen, Z., Morris, J. B., Norris, J. E., Barkley, N. A., Tonnis, B., Pederson, G. A. (2017). Genetic diversity and population structure of castor (*Ricinus communis* L.) germplasm within the US collection assessed with EST-SSR markers. *Genome* 60, 193-200. https://doi.org/10.1139/gen-2016-0116.

Xu, W., Yang, T., Qiu, L., Chapman, M.A., Li, D.Z., Liu, A. (2019). Genomic analysis reveals rich genetic variation and potential targets of selection during domestication of castor bean from perennial woody tree to annual semi-woody crop. *Plant Direct* 3, e00173. https://doi.org/10.1002/pld3.173.

Yol, E., Uzun, B. (2012). Geographical patterns of sesame accessions grown under Mediterranean environmental conditions, and establishment of a core collection. *Crop Sci.* 52, 1971-2421. https://doi.org/10.2135/cropsci2011.07.0355.

Yuan, Z., Rembe, M., Mascher, M., Stein, N., Jayakodi, M., Börner, A., Oldach, K., Jahoor, A., Jensen, J.D., Rudloff, J., Dohrendorf, V.-E., Kuhfus, L.P., Dyrszka, E., Conte, M., Hinz, F., Trouchaud, S., Reif, J.C., El Hanafi, S. (2024). Capitalizing on genebank core collections for rare and novel disease resistance loci to enhance barley resilience. *J. Exper. Bot.* 75, 5940-5954. https://doi.org/10.1093/jxb/erae283

Zangemeister-Wittke, U., Collinson, A.R., Fisch, I., Jones, R.M.L., Waibel, R., Lehman, H.-P., Stahel, R.A. (1993). Anti-tumor activity of a blocked ricin immunotoxin with specificity against the cluster-5A antigen associated with human small-cell lung cancer. *Internat. J. Cancer* 54, 1028-1035. https://doi.org/10.1002/ijc.2910540628.

Genetic Resources

ORIGINAL ARTICLE

Genetic Resources (2026), 7 (13), 192–201
DOI: 10.46265/genresj.WHMW8903
https://www.genresj.org
ISSN: 2708-3764

Phenomic characterization of *Crotalaria* germplasm in Embrapa's genebank, Brazil

João Marcelo Machado Squarisi[a], Mariane Rodrigues Ferreira[b], Juaci Vitoria Malaquias[b], Allan Kardec Braga Ramos[b], Claudio Takao Karia[b], Gustavo Jose Braga[b], Marcelo Ayres Carvalho[b,*]

a UniCeub, Campus da Asa Norte: SEPN 707/907 - Campus Universitário, Brazil
b Embrapa Cerrados, BR 020, km 18, Planaltina-DF, ZIP 73310-970, Brazil

Abstract: This study aimed to evaluate the morphological diversity of 22 *Crotalaria* accessions conserved in the Embrapa Cerrados germplasm bank, Brazil, with emphasis on their potential use in breeding and genetic resource management. Sixteen morphological descriptors were analyzed, revealing substantial phenotypic variation among accessions. Plant height ranged from 45 to 146cm, and leaflet and seed traits also showed broad variation, reflecting a wide adaptive spectrum. The Shannon–Weaver (H) and Simpson (D) diversity indices indicated consistently high diversity across traits (mean H = 0.964; mean D = 0.940), confirming the existence of a rich genetic base within the collection. Principal component analysis showed that the first two components explained 63.72% of total variance, with reproductive traits such as seed length, seed width, and 100-seed weight contributing most to variation among accessions. Cluster analysis grouped the accessions into five distinct morphological clusters, identifying genotypes with superior plant vigour, leaf area, or seed characteristics as potential parents for future breeding programmes. The observed morphological variability highlights the significant genetic diversity within Crotalaria germplasm, supporting its value for selection, hybridization and conservation. The combination of diversity indices and multivariate analyses provides a robust framework for identifying promising accessions for breeding programmes targeting biomass production, seed yield and environmental adaptation. These findings reinforce the importance of maintaining and characterizing germplasm collections as strategic reservoirs for sustainable crop improvement and the long-term conservation of genetic resources.

Keywords: Tropical legume, genetic resources, genetic diversity, morphology, cluster analysis

Citation: Squarisi, J. M. M., Ferreira, M. R., Malaquias, J. V., Ramos, A. K. B., Karia, C. T., Braga, G. J. and Carvalho, M. A. (2026) "Phenomic characterization of Crotalaria germplasm in Embrapa's genebank, Brazil", *Genetic Resources*, 7(13), pp. 192–201. doi: 10.46265/genresj.WHMW8903.

Introduction

The use of legumes in sustainable agricultural production systems has been widely recommended due to their ability to fix atmospheric nitrogen, produce high biomass, and promote physical, chemical and biological soil improvements. Among these legumes, the genus *Crotalaria* (Fabaceae) has stood out due to its hardiness, adaptation to tropical regions, and high potential for use as a cover crop and green manure (Daimon, 2006).

Studies by Koudahe *et al* (2022), Silva *et al* (2019) and Arone *et al* (2024) highlight the strategic role of *Crotalaria* in agroecological systems focused on the sustainability of tropical soils, especially due to its efficiency in nutrient cycling, restoration of degraded areas, and ecological multifunctionality.

The genus *Crotalaria* comprises over 300 species, some of which are popularly known as 'rattlepod' due to the characteristic rattling sound their mature seeds make inside their dry pods (Polhill, 1968). These fast-growing plants are widely used in agricultural systems, performing various ecological functions, such as green manure through biological nitrogen fixation (BNF), carbon fixation, nutrient cycling, nematode control, and attracting pollinating insects (Muller-Salmann and Kotschi, 1994). Despite this agricultural potential, which can be utilized in a variety of ways, there is still little information available on the genetic diversity stored in biological collections of the species. These genetic resources are extremely important for the selection and development of new cultivars for agricultural use.

* Corresponding author: Marcelo Ayres Carvalho (marcelo.ayres@embrapa.br)

Received: 09.10.2025 | Accepted: 06.05.2026 | Published online: xx.06.2026

Crotalaria species are mostly annual herbaceous plants, with some perennial or shrubby species. They have erect stems, which can reach 0.5 to 3 metres in height, with variable branching and the presence of trichomes, giving the plant a pubescent appearance (National Research Council, 1979). The leaves are alternate, simple (as in *C. juncea*) or trifoliate, and vary in shape from lanceolate to elliptical (Devecchi *et al*, 2014).

The root system is taproot-like, with a deep, well-developed taproot, accompanied by a vigorous secondary root system (Bolleddu *et al*, 2023). This morphological characteristic favours nutrient absorption at depth, improves soil porosity, and contributes to soil decompaction (Lanna *et al*, 2021). The flowers are papilionate, usually yellow, and arranged on terminal branches. The fruits are legume-like, with several seeds that produce a characteristic sound when dry, giving rise to the common name 'rattlepods' (Muli *et al*, 2025).

The conservation and characterization of genetic resources, both *in situ* and *ex situ*, is strategic to ensure the functional resilience of the genus in the face of climate change and agricultural intensification. National and international germplasm banks, such as those of Embrapa Cerrados, the International Livestock Research Institute (ILRI), the International Center for Tropical Agriculture (CIAT) and the US Department of Agriculture (USDA), have played a relevant role in the collection, characterization and provision of *Crotalaria* accessions for use in research and development (Muli *et al*, 2022).

Genetic and morphological studies reveal significant intraspecific and interspecific variability between cultivated and native *Crotalaria* species, offering significant selection potential for breeding programmes aimed at developing new cultivars with pest resistance, greater biomass production, adaptation to low-fertility soils, and greater efficiency in biological nitrogen fixation (Muli *et al*, 2022; Odhoch *et al*, 2025).

Crotalaria species exhibit a high adaptability to the Brazilian Cerrado, particularly *C. ochroleuca*, *C. juncea*, *C. breviflora* and *C. spectabilis*, which have been extensively evaluated for biomass production, BNF, and their effects on soil fertility and biology (Silva *et al*, 2025). Dry matter production can range from 5 to 18t/ha, depending on soil and climate conditions and management practices (Abranches *et al* 2021; Lima Filho *et al*, 2023).

Studies conducted in the Cerrado region indicate that species such as *C. ochroleuca* and *C. juncea* are particularly efficient at accumulating nutrients: average N, P and K contents in the shoots can reach 25g/kg, 3.5g/kg, and 15g/kg, respectively (Silva *et al*, 2025). Furthermore, these legumes accumulate high amounts of fibres such as cellulose and lignin, contributing to the formation of stable organic matter in the soil (Guerra *et al*, 2007; Arone *et al*, 2024).

BNF is one of the most valued attributes of the *Crotalaria* genus, with reports of fixation exceeding 300kg N/ha in crops of *C. juncea and C. spectabilis* in association with efficient strains of *Bradyrhizobium* (Ferreira *et al*, 2021). This symbiosis significantly reduces the need for nitrogen fertilizers and improves nutrient use efficiency in integrated systems (Ferreira *et al*, 2021).

Furthermore, *Crotalaria* possesses bioactive compounds with allelopathic and nematicidal properties. Species such as *C. spectabilis* and *C. juncea* are used in the management of plant-parasitic nematodes, such as *Meloidogyne incognita* and *Pratylenchus brachyurus*, due to the production of alkaloids that are toxic to these organisms (Wang *et al*, 2002; Oka, 2010).

Despite the traditional focus on its use as a cover crop, there is growing interest in improving *Crotalaria* seed productivity, especially given its expanding use in sustainable agricultural systems. Average seed production ranges from 500 to 1,200kg/ha, depending on the species, planting density, and environmental conditions (da Silva *et al*, 2022). Research focused on genetic improvement and harvest management has sought to optimize the viability and vigour of the seeds produced (da Silva *et al*, 2022).

Characterizing and evaluating the genetic diversity of these materials is essential to promote the use of genetic resources stored in germplasm banks, in addition to guiding actions to expand the available genetic base. Breeding programmes and the development of new cultivars can utilize the information resulting from efforts to characterize genetic variability and identify superior genotypes, helping to meet the demands of current and future agriculture. In this context, the study aimed to evaluate the genetic variability among the *Crotalaria* accessions stored in Embrapa Cerrados Germplasm Bank using morphological descriptors.

Materials and methods

Twenty-two accessions of *Crotalaria* spp. in the Embrapa Cerrados Germplasm Bank, located at Planaltina/DF, Brazil, were evaluated to determine their phenomic diversity (Table 1).

Table 1. List of *Crotalaria* spp. accessions in the Germplasm Bank of Embrapa Cerrados. *, Brazilian germplasm system number.

Id Number	BRA Number*	Species
2	00158911-8	*Crotalaria mitosa*
3	00158912-6	*Crotalaria spectabilis* Roth
4	00158913-4	*Crotalaria* sp.
6	00158921-7	*Crotalaria anagyroides* Kunth
7	00158916-7	*Crotalaria* sp.
8	00158917-5	*Crotalaria* sp.
10	00158919-1	*Crotalaria* sp.
11	00158920-9	*Crotalaria* sp.
12	00158914-2	*Crotalaria* sp.
13	00158915-9	*Crotalaria retusa* L.
14	00158922-5	*Crotalaria spectabilis* Roth
16	00158925-8	*Crotalaria* sp.
19	00158928-2	*Crotalaria striata* DC.
20	00158929-0	*Crotalaria ochroleuca* G. Don
21	00158939-9	*Crotalaria incana* L.
22	00158930-8	*Crotalaria mucronata* Desv.
24	00158932-4	*Crotalaria spectabilis* Roth
25	00158933-2	*Crotalaria retusa* L.
26	00158934-0	*Crotalaria retusa* L.
28	00158936-5	*Crotalaria grantiana* Harv.
29	00158937-3	*Crotalaria paulina* Schrank
30	00158938-1	*Crotalaria pallida* Aiton

Seeds of each accession were sown in 60-cell plastic trays filled with a suitable substrate, remaining there for 90 days after germination. After this period, they were transplanted to the field at Embrapa Cerrados Research Center, in Planaltina/DF/Brazil (15°35'34,42"S e 47°43'53,41"W), in a clayey red latosol, in single-row plots containing seven plants per plot, spaced with 0,5m among plants and 1,5m between lines. After 90 days, the vegetative measures were performed and the seeds characteristics were collected according to the maturation of the seeds, from each accession, using 16 qualitative and quantitative descriptors as presented in Table 2.

A graduated ruler and precision callipers were used to accurately measure morphological data. Seed weight was measured using samples of 100 seeds, with five replicates per accession. Seed colour was classified according to the Royal Horticultural Society (RHS, 2015) colour chart, and seed shape was defined based on pre-existing morphological classes (heart and kidney).

Genetic variability among the accessions with respect to the morphological descriptors was examined by calculating Simpson's (1949) and Shannon and Weaver's (1949) diversity indices. These indices give a measure of phenotypic diversity and range from zero to one, where one represents great genetic diversity and zero the opposite or no genetic diversity. The indices correspond to the probability that two individuals randomly selected from a group of populations will have the same morphological feature. The formulas for calculating both indices are presented below.

- Shannon–Weaver (H):

$$H = -\sum_{i=1}^{n} p_i \ln(p_i)$$

- Simpson (D):

$$D = 1 - \sum_{i=1}^{n} {p_i}^2$$

where $i = 1$ to n, and p is the proportion of the total morphotypes made up of the i^{th} morphotype.

The data were organized in digital spreadsheets and subjected to descriptive and multivariate analyses using R software (R Core Team, 2024). Principal component analysis (PCA) and cluster analysis were performed using the Ward method (Ward, 1963), with the aim of observing the genetic variability among the accessions and separating them into similarity groups. Qualitative descriptors were not used for PCA due to their categorical nature, making them unsuitable for PCA analysis.

Results

A wide range of variation was observed among the evaluated morphological descriptors (Table 3), indicating substantial phenotypic diversity within the studied accessions. Plant height exhibited the largest variation, with a mean of 81.4cm and a broad range from 45 to 146cm (SD = 26.8), reflecting differences in overall plant vigour and architecture. Similarly, plant diameter varied considerably (20.5–99cm, mean = 63.6cm), suggesting diverse growth habits and canopy structures among genotypes.

Traits related to branching showed moderate variability. The main branch diameter ranged from 4.6 to 17cm, while main branch length varied between 9 and 61.5cm. The number of primary branches exhibited a relatively high dispersion, highlighting morphological heterogeneity in plant architecture.

Comparing the individual accessions, accession #25 (BGF 6494) had plants with the tallest average height (145cm) and largest average diameter (92cm), exhibiting an upright growth habit. The smallest accession #13 (BGF 6476) had an average height of 45cm and a diameter of 54cm, exhibiting a semi-upright growth habit. Both accessions are *Crotalaria retusa* L. (Table 4).

Table 2. List of evaluated morphological descriptors for *Crotalaria sp.*

Morphological descriptor	Code	Replications	Unit	Equipment used
Plant height	AP	7	cm	Graduated ruler
Plant diameter	DP	7	cm	Graduated ruler
Growth habit	HC	7	Categories	Visual evaluation
Main branch diameter	CM	7	mm	Digital callipers
Main branch length	CRP	7	mm	Digital callipers
Number of primary branches	NRP	7	Count	Manual counting
Central leaflet length	CFC	7	mm	Digital callipers
Central leaflet width	LFC	7	mm	Digital callipers
Central leaflet L/W ratio	RC_LFC	7	mm	Calculated
Petiole length	CP	7	mm	Digital callipers
Leaflet shape	LS	7	Categories	Visual evaluation
Seed length	CS	7	mm	Digital callipers
Seed width	WS	7	mm	Digital callipers
Seed shape	FDS	7	Categories	Visual evaluation
Seed colour	Srgb	7	Colour chart	Visual evaluation
100-seed weight	PCS	5	g	Precision scale

Table 3. Descriptive statistics of quantitative descriptors

Morphological descriptor	Mean	Standard deviation	Variance	Range
Plant height (cm)	81.4	26.8	718.3	45–146
Plant diameter (cm)	63.6	18.7	350.2	20.5–99
Main branch diameter (cm)	9.7	3.2	10.2	4.6–17
Main branch length (cm)	31.6	13.5	182.3	9–61.5
Number of primary branches	10.7	4.7	22.3	6–25
Central leaflet length (mm)	60.1	26.3	692.7	10.7–105
Central leaflet width (mm)	27.0	11.2	125.4	4.6–51
Central leaflet L/W ratio	2.6	1.5	2.3	1.47–8.4
Petiole length (mm)	20.9	15.3	233.9	2–44
Seed length (mm)	2.5	0.6	0.36	2–3.57
Seed width (mm)	3.3	0.9	0.87	2.16–4.68
100-seed weight (g)	1.2	0.7	0.52	0.40–2.19

Leaflet traits also displayed marked variation. Central leaflet length ranged from 10.7 to 105mm, and leaflet width from 4.6 to 51mm. The length-to-width ratio (L/W) varied from 1.47 to 8.4, suggesting the coexistence of both broad and elongated leaflet morphotypes. Petiole length ranged from 2 to 44mm, showing substantial variability in leaf structure.

Seed-related traits were comparatively less variable. Seed length and width ranged from 2 to 3.57mm and 2.16 to 4.68mm, respectively, with low standard deviations. The 100-seed weight exhibited moderate variation, consistent with the general trend of smaller seed-size diversity compared with vegetative descriptors.

Regarding leaflet shape, it was observed that ten accessions had trifoliate leaves, 11 had simple leaves, and only one had a digitate leaf, demonstrating significant morphological variation in this characteristic (Figure 1). In addition, significant differences in shape and colour of the seeds were observed. The majority of the accessions (18) presented a kidney shape, and the remaining accessions (4) presented a heart shape seed (Figure 1). In relation to the colour of the seeds, six different types were observed: yellow (6), green (6), orange (2), black (4), red (1), and striped (1) were observed (Figure 1).

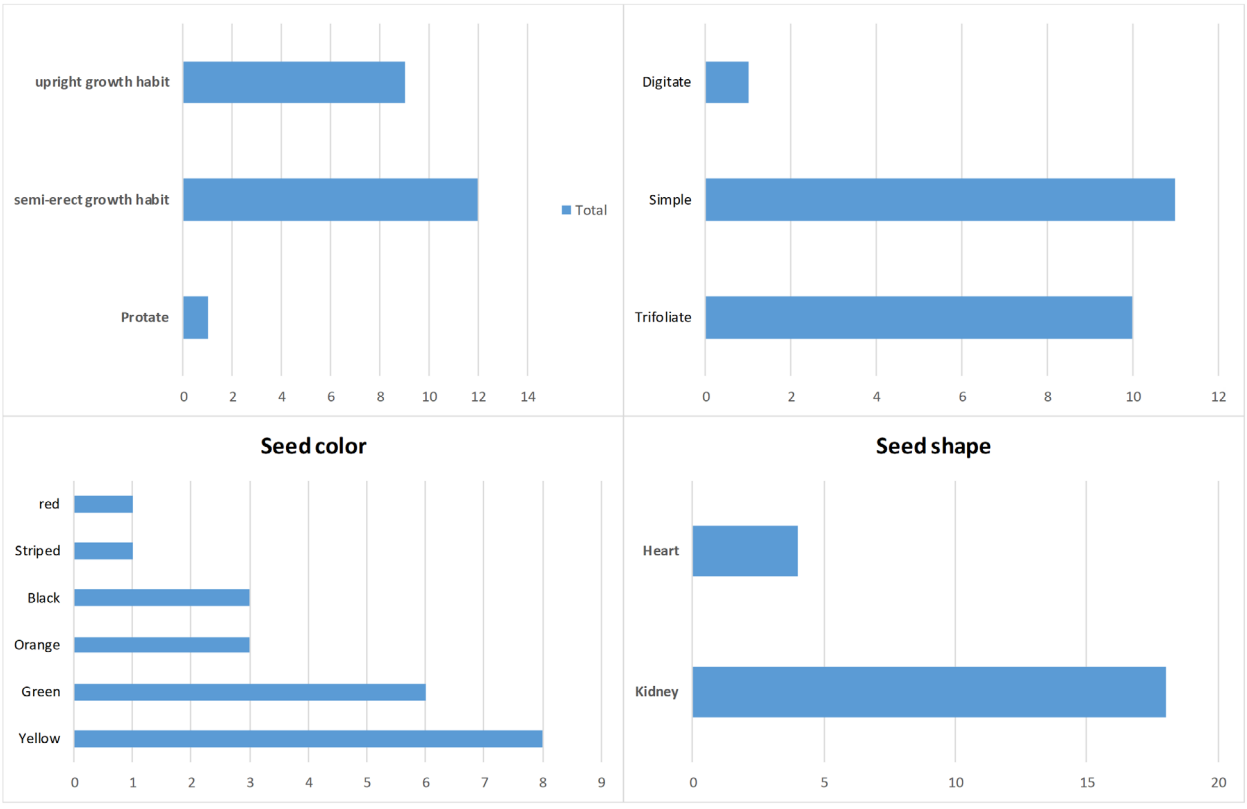


Figure 1. Frequency distribution of growth habit, leaflet shape, seed colour and shape observed on the *Crotalaria* germplasm.

Table 4. Plant height (AP); plant diameter (DP); growth habit (HC); main branch diameter (DHP); main branch length (CHP); number of primary branches (NRP); central leaflet length (CFC); central leaflet width (LFC); petiole length (CP); leaflet shape (LS); seed length (CS); seed width (WS); Seed shape (FDS); Seed colour (Srgb) and 100-seed weight (PCS) of Crotalaria accessions.

ID #	AP (cm)	DP (cm)	HC	CM (mm)	CRP (cm)	NRP	CFC (mm)	LFC (mm)	RC_LFC (cm)	CP (mm)	LS	CS (mm)	WS (mm)	FDS	Srgb	PCS (g)
2	78	62	Upright	11	30	6	60.75	34	1.81	31	trifoliate	2.3	2.18	heart	yellow	0.532
3	83	63	Upright	9	26	7	43.4	24.9	1.76	33.2	trifoliate	2.15	2.8	kidney	green	0.693
4	69	60	Upright	6	25	7	65	36	1.8	35	trifoliate	2.18	2.88	kidney	green	0.664
6	93	64	Upright	12	22.5	10	39	20.5	1.92	31	trifoliate	2.65	2.47	heart	yellow	0.661
7	59	68	Semi-erect	7	42.2	9	47.5	18.7	2.52	2.7	simple	3.19	4.2	kidney	yellow	1.88
8	53	51	Semi-erect	6.5	42.5	10	50.3	16.4	3.1	2.3	simple	3.28	4	kidney	yellow	1.77
10	102	75	Semi-erect	6.3	61.5	9	85	3.5	26.8	2	simple	2.5	2.79	kidney	black	0.760
11	54	49	Semi-erect	6	41	8	48.8	17.5	2.81	2.1	simple	3.28	4.13	kidney	yellow	1.83
12	58	99	Prostate	10	38	15	42.5	17.5	2.5	44	trifoliate	2.37	3	kidney	orange	0.777
13	45	54	Semi-erect	7	48	8	58.2	20.7	2.88	2.4	simple	3.36	4.35	kidney	yellow	2.09
14	99	51	Semi-erect	13	28	19	105	42.5	2.5	4.5	simple	3.57	4.61	kidney	black	2.19
16	67	53	Upright	12	32	11	98.9	40.2	2.48	4.8	simple	3.01	3.68	kidney	red	1.08
19	115	83	Upright	16	16	8	48.1	32.8	1.47	28.5	trifoliate	3.08	2.7	heart	green	0.878
20	146	92	Upright	17	9	25	57.5	7.75	8.4	31.2	trifoliate	2	3,31	kidney	orange	1.13
21	115	76	Upright	7.6	14	12	82	29	2.8	8	simple	2.4	2.87	kidney	green	0.719
22	67	20.5	Semi-erect	4.6	18	10	46	21.7	2.22	36	trifoliate	2.48	2.85	kidney	striped	0.766
24	63	36	Semi-erect	8.8	22	6	103.7	51	2	5.6	simple	3.57	4.68	kidney	black	2.04
25	52	44	Semi-erect	7	26.5	9	48	16.8	2.9	2.9	simple	3.27	4.42	kidney	yellow	1.96
26	54	47	Semi-erect	6	35	7	43.2	14.8	2.9	2.5	simple	3.41	4.31	kidney	yellow	2.1
28	70	46	Semi-erect	7.8	19.5	9	10.7	4.6	2.5	5.4	digitate	2	2.16	kidney	orange	0.4
29	82	84	Upright	11	28	6	41.6	21	1.97	39.4	trifoliate	2.34	2.54	kidney	green	0.583
30	90	80	Semi-erect	10.5	24.6	8	53.5	24.7	2.1	38.6	trifoliate	2.62	2.43	heart	green	0.634

Table 5. Shannon–Weaver (H) and Simpson (D) diversity indices for quantitative descriptors

Descriptor	Shannon–Weaver (H)	Simpson (D)
Plant height (cm)	0.984	0.950
Plant diameter (cm)	0.985	0.950
Main branch diameter (cm)	0.981	0.949
Main branch length (cm)	0.974	0.947
Number of primary branches	0.974	0.946
Central leaflet length (mm)	0.975	0.948
Central leaflet width (mm)	0.966	0.945
Central leaflet L/W ratio	0.971	0.943
Petiole length (mm)	0.865	0.920
Seed length (mm)	0.990	0.952
Seed width (mm)	0.942	0.939
100-seed weight (g)	0.956	0.942
Total	0.964	0.940

Shannon–Weaver (H) and Simpson (D) indices were calculated using the continuous values of quantitative morphological descriptors, treating each accession as a unique morphotype (Table 5).

Diversity values were consistently high across all evaluated traits. Shannon accurately represents class richness and the presence of rare classes, showing values closer to 1 when morphologically unique accessions are present, even if at low frequency. Simpson, on the other hand, is guided by the concentration in a few classes (dominance). The value is high when many accessions are very closely related/duplicated.

The Shannon–Weaver index (H) ranged from 0.865 for petiole length to 0.990 for seed length, with a total mean of 0.964. Similarly, Simpson's index (D) varied from 0.920 for petiole length to 0.952 for seed length, with a total mean of 0.940.

Among vegetative traits, plant height (H = 0.984; D = 0.950), plant diameter (H = 0.985; D = 0.950), and main branch diameter (H = 0.981; D = 0.949) showed the highest diversity values. Leaflet traits displayed slightly lower but still high values, with central leaflet length (H = 0.975; D = 0.948) and width (H = 0.966; D = 0.945) presenting similar diversity patterns.

Seed traits exhibited high and uniform diversity levels. Seed length had the highest indices among all descriptors (H = 0.990; D = 0.952), followed by seed width (H = 0.942; D = 0.939) and 100-seed weight (H = 0.956; D = 0.942). Overall, both indices indicated substantial phenotypic variability among the evaluated quantitative traits.

PCA was performed to discriminate among accessions and group them into different clusters/groups. The goal of PCA is to provide a reduced dimension model that would indicate measured differences among groups. It can also contribute to a better understanding of the set of variables by describing how much of the total variance is explained by each one. With this objective, PCA was performed only on quantitative variables. Thus, only 11 variables were used: plant height (AP); plant diameter (DP); main branch diameter (DHP); main branch length (CHP); number of primary branches (NRP); central leaflet length (CFC); central leaflet width (LFC); petiole length (CP); seed length (CS); seed width (WS); and 100-seed weight (PCS).

The PCA revealed that the first two components explain 63.72% of the total data variation, which indicates a good representation of the genetic variability among the evaluated accessions (Table 6).

Table 6. Vector loadings and percentage of variation explained by the first three principal components for morphological characteristics of *Crotalaria sp.*

Variable	**PC1**	**PC2**	**PC3**
Eigen Value	4.56	2.45	1.47
% Explained	41.48	22.24	13.37

The loadings of morphological descriptors on each component indicated that PC1 was primarily influenced by seed length, seed width, and 100-seed weight (Table 7). PC2 was formed based on main branch diameter (CM), central leaflet length (CFC) and number of primary branches (NRP).

Table 7. Loadings of morphological descriptors on principal components. AP (plant height); DP (plant diameter); CM (main branch diameter); CRP (main branch length); NRP (number of primary branches); CFC (central leaflet length); LFC (central leaflet width); CP (petiole length); CS (seed length); WS (seed width); PCS (100-seed weight).

Descriptor	**PC1**	**PC2**
AP	-0.686	0.555
DP	-0.651	0.194
CM	-0.538	0.692
CRP	0.517	-0.357
NRP	-0.269	0.632
CFC	0.329	0.699
LFC	0.199	0.541
CP	-0.788	-0.079
CS	0.868	0.295
WS	0.871	0.356
PCS	0.872	0.306

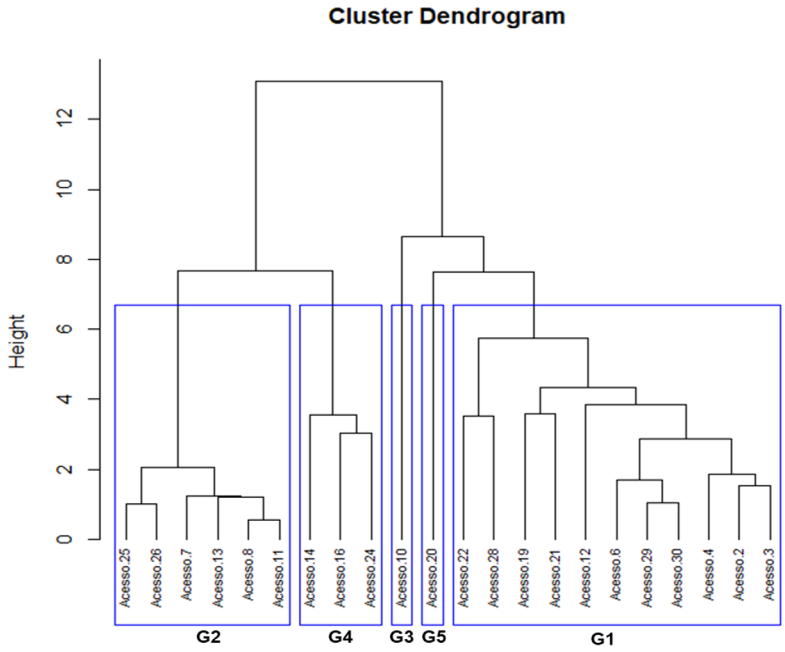


Figure 2. Dendrogram of the 22 *Crotalaria* accessions based on morphological descriptors using the Ward method

The first two principal components were used to execute a cluster analysis using the complete linkage clusters method (Sokal and Michener, 1958). The dendrogram resulting from this analysis is presented in Figure 2.

A cluster analysis using the Ward method allowed the 22 accessions to be divided into five groups based on similarity. Group 1 was the largest, with 11 accessions, followed by Group 2 with six accessions, and Group 4 with three accessions. Groups 3 and 5 each had only one accession; these accessions correspond to the tallest and largest plants (Table 8).

Group 1 was represented by nine different species, demonstrating that it included the largest diversity. Group 2 was mostly represented by *C. retusa* accessions. Group 3 was formed by a single *Crotalaria sp.* accession. Group 4 gathered mostly accessions of *C. spectabilis* and Group 5 was formed by a single accession of *C. ochroleuca*.

Discussion

Phenotypic variation

The morphological evaluation of the accessions revealed substantial variability across all traits, indicating a high degree of phenotypic diversity. Most of the accessions presented an upright (11 accessions) or a semi-erect growth habit (10). This variation is critical for selection depending on the intended agronomic purpose, such as forage production or adaptation to specific environments (Sayed *et al*, 2022). Similarly, plant diameter (DP) showed moderate variation, reflecting differences in plant architecture that may influence light interception and biomass accumulation (Solbrig, 1994). Accessions with greater height and diameter may be prioritized in breeding programmes aiming to maximize biomass, while more compact forms could be selected for environments requiring denser planting or easier management.

Branch traits, including main branch diameter (DHP) and main branch length (CHP), exhibited considerable diversity. Such variation affects structural stability and biomass yield, particularly in accessions with higher numbers of branches (NRP), which ranged from 6 to 25. These branching traits are important for breeding strategies targeting canopy architecture optimization, light distribution and overall productivity (Nelson and Moser, 1994).

Leaf size influences photosynthetic capacity, transpiration and overall plant growth, contributing to adaptation under different environmental conditions. Accessions with larger leaf areas could be prioritized for breeding programmes focused on maximizing biomass accumulation, while smaller leaves may be advantageous for water-limited environments (Weraduwage *et al*, 2015). Accessions #2, 4, 14, 16, 19, 21

Table 8. Averages of morphological descriptors by group AP (plant height); DP (plant diameter); CM (main branch diameter); CRP (main branch length); NRP (number of primary branches); CFC (central leaflet length); LFC (central leaflet width); CP (petiole length); CS (seed length); WS (seed width); PCS (100-seed weight).

Cluster	Nº	AP	DP	CM	CRP	NRP	CFC	LFC	CP	WS	CS	PCS
G1	11	83.6	67.0	9.6	23.8	8.9	48.4	24.2	30.0	2.4	2.6	0.7
G2	6	52.8	52.6	8.7	16.0	5.7	42.5	17.5	25.0	3.3	3.4	2.1
G3	1	102.0	75.0	6.1	18.0	9.0	55.0	34.5	20.0	2.5	2.8	0.8
G4	3	80.0	72.0	17.0	40.0	5.0	65.0	30.0	14.0	2.0	2.0	1.0
G5	1	146.0	92.0	17.0	90.0	25.0	57.5	7.8	31.2	2.0	3.3	1.1

and 24 presented the larger leaf area among the germplasm.

Seed weight (PCS) ranged from 0.40 to 2.19g, indicating potential differences in seedling vigour and establishment. Twelve of the 24 accessions (half) had a seed weight below 1g, while the other half exceeded 1g. The presence of approximately seven accessions classified only as *Crotalaria* sp. within these two groups complicates the determination of whether this difference is genuinely species-related. Nevertheless, there are three accessions classified as *C. spectabilis*. The PCS for these accessions presented large variability: 0,7g; 2g and 2.2g. Variation in seed weight is a critical trait for germplasm evaluation and breeding, as it often correlates with seedling performance, establishment success, and yield potential (Gnan *et al*, 2014). Accessions with higher seed weights may be selected to enhance early growth and establishment in breeding programmes targeting improved crop performance.

The observed diversity in seed shape and colour, as well as in leaf morphology (trifoliate, simple, and digitated), confirms significant genetic variability and underscores the importance of *ex situ* conservation in the Embrapa Cerrados germplasm bank. The presence of accessions with unique characteristics, such as heart- or kidney-shaped seeds and distinct growth habits, highlights the availability of genetic resources for breeding programmes targeting seed production, biomass yield and nutrient-use efficiency (Arone *et al*, 2024; Muli *et al*, 2021).

Overall, the observed morphological variation suggests a rich genetic base within the evaluated accessions. Vegetative traits, particularly plant height, leaflet size and branching characteristics, exhibited higher coefficients of variation than reproductive traits, emphasizing their contribution to morphological differentiation among accessions. Integrating these phenotypic evaluations into germplasm selection strategies can facilitate the identification of superior accessions, providing a foundation for efficient breeding programmes. This broad genetic diversity could also be exploited in breeding programmes targeting plant architecture and leaf morphology, resulting in new cultivars with improved adaptability, biomass productivity, or specific agronomic characteristics. Future studies combining molecular marker analysis with these morphological traits would provide a more comprehensive understanding of the genetic potential and heritability of key traits.

Diversity index

The analysis of morphological traits in this study provides valuable insights into the genetic diversity of the evaluated accessions. The Shannon–Weaver (H) and Simpson (D) diversity indices offer a quantitative assessment of this diversity, which is crucial for effective breeding and germplasm selection strategies.

The quantitative morphological descriptors showed an average Shannon–Weaver diversity index (H) of 0.964 and an average Simpson diversity index (D) of 0.940. When H and D values are high and close to 1 in a germplasm collection calculated from morphological descriptors, this generally indicates high phenotypic diversity and low dominance of a morphological state. Traits such as AP (H = 0.984), DP (H = 0.985), and CS (H = 0.990) exhibited high diversity, indicating a broad genetic base. Conversely, traits like NRP (H = 0.65) showed lower diversity, suggesting potential areas for improvement through breeding.

A high H value indicates many states present (class richness) and/or a relatively balanced distribution among these states (evenness). In practice, this means that the collection covers a wide range of phenotypes and is not concentrated in a few morphological types.

A high D value (approaching 1) indicates the absence of a clearly dominant morphological state among the analyzed descriptors. This suggests that, based on these descriptors, the germplasm bank contains few 'identical' accessions, resulting in lower redundancy (fewer evident phenotypic duplicates).

These results indicate that the collection exhibits good stratification and represents a wide variation, which is valuable for *ex situ* conservation. The existing diversity, already well distributed, increases the potential for assembling core collections with minimal loss of variability, while the presence of extreme or rare phenotypes provides additional options for pre-breeding.

Accessions with unique trait profiles, especially those with lower diversity, can serve as valuable sources of novel alleles. Incorporating these into breeding programmes can broaden the genetic base and introduce beneficial traits (Fu *et al*, 2015).

The diversity indices can inform the development of selection indices that prioritize traits with optimal diversity levels, balancing the need for improvement with the preservation of genetic variability.

Diversity assessments in other legume species have demonstrated that combining diversity indices such as Shannon–Weaver and Simpson with multivariate analyses is an effective approach for characterizing germplasm collections and guiding breeding and conservation strategies (Upadhyaya *et al*, 2002; Carvalho and Quesenberry, 2009).

Principal component analysis

PC1 was primarily influenced by reproductive characteristics, specifically seed length, seed width, and 100-seed weight (Table 6). These positive loadings suggest that seed size and weight were the main sources of variation among the accessions. Conversely, a morphological trade-off was indicated by the negative correlations of PC1 with vegetative traits, such as petiole length (-0.788), plant height (-0.686), and plant diameter (-0.651). This pattern – a trade-off between vegetative growth and seed size – is consistent with findings reported by Devecchi *et al* (2014), Lanna *et al* (2021) and Carvalho and Quesenberry (2009).

PC2 was predominantly associated with plant height (AP), main branch diameter (CM), number of primary branches (NRP), central leaflet length (CFC) and central leaflet width (LFC).

Cluster analysis

Group 1 (11 accessions) represents the 'mainstream' ideotype: medium plant height and canopy diameter, moderate stem and primary-branch development, intermediate leaflet and petiole dimensions and the smallest seeds with the lowest 100-seed weight. Agronomically, this profile fits a general-purpose cover/green-manure pool with reasonably balanced architecture. Group 2 (6 accessions) is a compact pool but clearly stands out for larger, heavier seeds, which is typically associated with better field emergence

and seed-lot robustness – useful for seed production and for service-crop cultivars where establishment reliability is a priority, including intercropping scenarios where lower stature reduces competition. Group 3 (1 accession) and Group 4 (3 accessions) concentrate 'leaf-and-structure' extremes: Group 3 has tall and wide plants with large central leaflets and moderate branching, suggesting strong shading potential for weed suppression and fast canopy closure (Abranches *et al*, 2021). Group 4 combines medium height with thick stems and very large leaflets but fewer primary branches, pointing to structural robustness and a broad-leaf canopy that can be valuable for biomass/canopy design depending on lodging behaviour. Group 5 (1 accession) is the most extreme architecture ideotype – very tall, very wide, exceptionally branched (#25) with very long main-branch length – indicating a high-biomass, high-competitiveness pool for aggressive cover, soil protection, and rehabilitation, while its comparatively heavier seeds support establishment.

These findings are consistent with previous studies on *Crotalaria* species, where PCA and cluster analyses have been employed to assess morphological diversity and accessions grouping based on phenotypic traits (Muli *et al*, 2021; Yaradua *et al*, 2018). The combination of PCA and cluster analysis in this study provides a comprehensive understanding of the morphological variation among *Crotalaria* accessions, highlighting the potential for selection and breeding programmes aimed at improving agronomic performance.

The comprehensive morphological evaluation of the *Crotalaria* accessions revealed a high degree of phenotypic and genetic diversity across vegetative and reproductive traits. The wide variation observed in plant height, diameter, branching pattern, leaflet morphology and seed characteristics underscores the existence of a rich and heterogeneous genetic base within the studied germplasm. The consistently high Shannon–Weaver and Simpson diversity indices confirm that the collection maintains a broad spectrum of variability, which is essential for effective germplasm conservation and future breeding initiatives.

The principal component and cluster analyses provided a clear separation of accessions into distinct morphological groups, indicating that the evaluated traits successfully captured meaningful genetic differentiation. Accessions exhibiting superior plant vigour, leaf area, or seed size were identified as potential parents for breeding programmes targeting biomass accumulation, seed production, or adaptability to diverse environments. Meanwhile, genotypes with contrasting phenotypes represent valuable resources for introgressing novel alleles and maintaining broad genetic variability within the species.

Conclusion

The results of this study have practical implications for both breeding and germplasm conservation. The identification of distinct phenotypic groups and accessions with superior agronomic traits provides a solid foundation for developing improved cultivars with enhanced biomass production, seed yield and environmental adaptability. Accessions representing divergent morphological clusters should be prioritized in hybridization programmes to maximize heterosis and maintain broad genetic variability. Furthermore, the high within-collection diversity observed reinforces the importance of continuous phenotypic and molecular characterization in germplasm banks to ensure the long-term preservation of genetic resources and their effective utilization in sustainable agricultural systems.

In summary, the study demonstrates that morphological characterization, combined with diversity indices and multivariate analyses, is a powerful approach for assessing genetic variability in *Crotalaria*. The identified diversity provides a solid foundation for future breeding efforts aimed at optimizing plant architecture, seed yield, and adaptation, while supporting the conservation and strategic utilization of genetic resources essential for crop improvement and ecosystem sustainability.

Author contributions

JMMS conducted data collection and manuscript writing. JVM performed the data analysis. AKBR, CTK, MRF and GJB contributed to manuscript review. MAC served as germplasm curator and contributed to experimental planning, data analysis and manuscript writing.

Conflict of interest statement

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The authors declare that this study received funding from Embrapa. The funder was not involved in the study design, collection, analysis, interpretation of data, the writing of this article or the decision to submit it for publication.

References

Abranches, M. de O., da Silva, G. A. M., dos Santos, L. C., Pereira, L. F., de Freitas, G. B. (2021). Contribution of green fertilization to the chemical, physical and biological characteristics of the soil and its influence on the nutrition of vegetables. *Research, Society and Development* 10, e3210716351. doi: https://doi.org/10.33448/rsd-v10i7.16351

Arone, G. J., Ocaña, R., Sánchez, A., Villadas, P. J., Fernández-López, M. (2024). Benefits of *Crotalaria juncea* L. as green manure in fertility and soil microorganisms on the Peruvian coast. *Microorganisms* 12, 2241. doi: https://doi.org/10.3390/microorganisms12112241

Bolleddu, R., Debajyoti, D., Das, M., Sahu, D., Mangal, A., Babu, G., Venkata, P., Prasad, P. (2023). Pharmacognostical and phytochemical studies of *Crotalaria retusa* L. (Shanapuspi) dried roots. *Journal of Drug Research in Ayurvedic Sciences* 8, 370–376. doi: https://doi.org/10.4103/jdras.jdras_31_23

Carvalho, M. A., Quesenberry, K. H. (2009). Morphological characterization of the USA Arachis pintoi Krap. and Greg. collection. *Plant Systematics and Evolution* 277, 1–11. doi: https://doi.org/10.1007/s00606-008-0089-9

da Silva, B. N. P., Masetto, T. E., Garcia, R. A., Toledo, M. Z. (2022). Row spacing and seed physiological quality of *Crotalaria* species. *Pesq. Agropec. Trop*. 52, e72674. doi: https://doi.org/10.1590/1983-40632022v5272674

Daimon, H. (2006). Traits of the Genus *Crotalaria* Used as a Green Manure Legume on Sustainable Cropping Systems. *JARQ* 40, 299–305. doi: https://doi.org/10.6090/jarq.40.299

Devecchi, M. F., Pirani, J. R., Melo-de-Pinna, G. F. d. A. (2014). Comparative leaf anatomy and morphology of some Brazilian species of *Crotalaria* L. (Leguminosae: Papilionoideae: Crotalarieae). *Acta Botanica Brasilica* 28, 583–593. doi: https://doi.org/10.1590/0102-33062014abb3517

Ferreira, E. A., et al (2021). Fixação biológica de nitrogênio em leguminosas de cobertura. *Cadernos de Agroecologia* 16, 2. doi: https://doi.org/10.22308/cadernosdeagroecologia.v16i2.17256

Fu, Y. B., et al (2015). Understanding crop genetic diversity under modern plant breeding. *Theor Appl Genet* 128, 2131–2142. https://doi.org/10.1007/s00122-015-2585-y

Gnan, S., Priest, A. and Kover, P.X., 2014. The genetic basis of natural variation in seed size and seed number and their trade-off using *Arabidopsis thaliana* MAGIC lines. Genetics, 198(4), pp.1751-1758. doi: https://doi.org/10.1534/genetics.114.170746

Guerra, J. G. M., Espíndola, J. A. A., Perin, A., Teixeira, M. G., Almeida, D. L. de, Assis, R. L. de (2007). Desempenho de leguminosas tropicais perenes como plantas de cobertura do solo. *Embrapa Agrobiologia, Boletim de Pesquisa e Desenvolvimento* 20. url: https://www.infoteca.cnptia.embrapa.br/infoteca/handle/doc/622591

Koudahe, K., Allen, S. C., Djaman, K. (2022). Critical review of the impact of cover crops on soil properties. *International Soil and Water Conservation Research* 10, 343–354. doi: https://doi.org/10.1016/j.iswcr.2022.03.003

Lanna, A. C., Silva, M. A., Moreira, A. S., Nascente, A. S., Filippi, M. C. C. de (2021). Improved nutrient uptake in three *Crotalaria* species inoculated with multifunctional microorganisms. *Rev. Bras. Eng. Agríc. Ambient*. 25, 460–465. doi: https://doi.org/10.1590/1807-1929/agriambi.v25n7p460-465

Lima Filho, O. F., et al, eds. (2023). Adubação verde e plantas de cobertura no Brasil: fundamentos e prática (Brasília, DF: Embrapa), 586 p.

Muli, J. K., Neondo, J.O., Kamau, P.K. et al. Phenomic characterization of Crotalaria germplasm for crop improvement. *CABI Agric Biosci* 2, 10 (2021). https://doi.org/10.1186/s43170-021-00031-0

Muli, J. K., Neondo, J. O., Kamau, P. K., Michuki, G. N., Odari, E., Budambula, N. L. M. (2022). Genetic diversity and population structure of wild and cultivated *Crotalaria* species based on genotyping-by-sequencing. *PLOS ONE* 17, e0272955. doi: https://doi.org/10.1371/journal.pone.0272955

Muli, J. K., Neondo, J. O., Budambula, N. L. M. (2025). Agro-morphological diversity of edible Kenyan *Crotalaria* accessions. *Scientific African* 27, e02620. doi: https://doi.org/10.1016/j.sciaf.2025.e02620

Müller-Samann, K. M., Kotschi, J. (1994). Sustaining growth: soil fertility management in tropical small holdings (United Kingdom: Technical Centre for Agricultural and Rural Cooperation), 486 p.

National Research Council. (1979). Tropical legumes: resources for the future (Washington, D.C.: National Academy of Science), 331 p.

Nelson, C. J., Moser, L. E. (1994). Plant factors affecting forage quality. In Forage Quality, Evaluation, and Utilization, ed. G. C. Fahey Jr., M. Collins, D. R. Mertens, L. E. Moser (Madison, WI: ASA, CSSA, SSSA), 115–154.

Odhoch, P. S., Budambula, N. L. M., Kiprotich, F., Muli, J. K. (2025). Exploring genetic diversity: Optimizing simple sequence repeat (SSR) markers in *Crotalaria* for enhanced precision in biodiversity research. Scientifica (Cairo) 2025, 2409286. doi: https://doi.org/10.1155/sci5/2409286

Oka, Y. (2010). Mechanisms of nematode suppression by organic soil amendments—A review. *Appl. Soil Ecol.* 44, 101–115. doi: https://doi.org/10.1016/j.apsoil.2009.11.003

Polhill, R. M. (1968). Miscellaneous notes on African species of *Crotalaria* L. II. *Kew Bulletin* 22, 169–337. doi: https://doi.org/10.2307/4107762

R Core Team. (2024) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. https://www.R-project.org/

Royal Horticultural Society (2015). RHS Colour Chart (6th ed.). London: Royal Horticultural Society. [Plate 155D]

Sayed, M. R. I., et al (2022). Genetic diversity, analysis of some agro-morphological and quality traits and utilization of plant resources of alfalfa. *Genes* 13, 1521. doi: https://doi.org/10.3390/genes13091521

Silva, J. L., Almeida, J., Sousa, M., Silva, E. (2018). Crotalária: usos e potencialidades no Cerrado. In Anais do III Congresso Internacional das Ciências Agrárias (COINTER PDVAgro), João Pessoa. doi: https://doi.org/10.31692/2526-7701.XCOINTERPDVAgro.2018.00033

Shannon, C.E., Weaver, W., 1949. The Mathematical Theory Communication. University of Illinois Press, Urbana.

Simpson, E. H. (1949). Measurement of diversity. *Nature* 163, 688. doi: https://doi.org/10.1038/163688a0

Solbrig, O. T. (1994). Plant traits and adaptive strategies: their role in ecosystem function. In Biodiversity and Ecosystem Function, ed. E.-D. Schulze, H. A. Mooney (Berlin: Springer), 97–116.

Sokal, R. R., & Michener, C. D. (1958). A statistical method for evaluating systematic relationships. *University of Kansas Science Bulletin*, 38, 1409-1438. [1, 2, 3]

Upadhyaya, H. D., Bramel, P.J., Ortiz, R. et al. (2002) Geographical patterns of diversity for morphological and agronomic traits in the groundnut germplasm collection. *Euphytica* 128, 191–204. https://doi.org/10.1023/A:1020835419262

Wang, K.-H., Sipes, B. S., Schmitt, D. P. (2002). *Crotalaria* as a cover crop for nematode management: a review. *Nematropica* 32, 35–58. url: https://journals.flvc.org/nematropica/article/view/69643

Ward, J. H., Jr. (1963). "Hierarchical grouping to optimize an objective function." *Journal of the American Statistical Association*, 58(301), 236-244.

Weraduwage, S. M., Chen, J., Anozie, F. C., Morales, A., Weise, S. E., Sharkey, T. D. (2015). The relationship between leaf area growth and biomass accumulation in *Arabidopsis thaliana*. *Frontiers in Plant Science* 6, 167. doi: https://doi.org/10.3389/fpls.2015.00167

Yaradua, S. S., et al (2018). Numerical taxonomic study of the genus *Crotalaria* L. (Crotalarieae, Fabaceae) in Nigeria. *Biodiversity Research and Conservation,* vol. 50, no. 1, Adam Mickiewicz University, pp. 25-32. https://doi.org/10.2478/biorc-2018-0005