Tanwar *et al. BMC Plant Biology* (2025) 25:1256
https://doi.org/10.1186/s12870-025-07115-3

BMC Plant Biology

RESEARCH

Open Access

Establishment and phenotypic characterization of genetically diverse core collections of *Lupinus albus*

Umesh Kumar Tanwar[1], Magdalena Tomaszewska[1], Katarzyna Czepiel[1], Mohamed Neji[1], Humaira Jamil[1], Lorenzo Rocchetti[2], Alice Pieri[2], Elena Bitocchi[2], Elisa Bellucci[2], Barbara Pipan[3], Vladimir Meglič[3], Markus Oppermann[4], Magdalena Kroc[1], Roberto Papa[2] and Karolina Susek[1*]

Abstract

Background *Lupinus albus* is a food grain legume recognized for its high levels of seed protein (30–40%) and oil (6–13%), and its adaptability to different climatic and soil conditions. The availability of well-characterized, genetically and phenotypically diverse germplasm will facilitate the development of next-generation *L. albus* cultivars, encourage biodiversity conservation, and promote the sustainable utilization of this species.

Results We evaluated more than 2000 *L. albus* accessions based on 35 agro-morphological traits and passport data to establish *Intelligent Collections*. The Reference-CORE (R-CORE), covering global diversity, exemplified the genotypic variation among accessions differing in biological status (cultivars, breeding/research materials, landraces, and wild relatives). The Training-CORE (T-CORE), a subset of 300 R-CORE accessions, represents the diversity of the entire collection. Principal component analysis showed that the *L. albus* R-CORE encompasses four phenotypic groups (A1, A2, A3, and B), and that groups A3 and B can be characterized by the main phenotypic traits of pod shattering and seed ornamentation, respectively. The coefficient of total genetic variation differed across morphological traits, phenotypic groups, geographic regions, and according to biological status.

Conclusions The core collections established in this study will facilitate agricultural research by providing the broad phenotypic data needed for crop improvement programs and by shedding light on the undiscovered biodiversity of *L. albus* genetic resources. Understanding the variation found in such resources will allow us to develop sustainable tools and technologies that address global challenges such as the provision of healthy and sustainable diets for all and the mitigation of climate change.

Keywords *Lupinus albus*, Plant genetic resources, Phenotypic diversity, Germplasm conservation, Core collection, Food legumes

---

*Correspondence:
Karolina Susek
ksus@igr.poznan.pl
[1]Legume Genomics Team, Institute of Plant Genetics, Polish Academy of Sciences, Strzeszynska 34, Poznan 60479, Poland
[2]Department of Agricultural, Food and Environmental Sciences, Università Politecnica delle Marche, Ancona 60131, Italy
[3]Crop Science Department, Agricultural Institute of Slovenia, Hacquetova ulica 17, Ljubljana SI-1000, Slovenia
[4]Research Group Genebank Documentation, Genebank Department, Institute of Plant Genetics and Crop Plant Research, Corrensstr. 3, 06466 Seeland OT Gatersleben, Germany

Background

Plant genetic resources (PGRs) are needed to preserve the genetic diversity that underpins the resilience, productivity, and sustainability of agricultural systems. The exploration and utilization of PGRs is hindered by accession heterogeneity, the lack of comprehensive passport data and descriptors valuable for users, and limited access to available information [1]. This impedes the integration of PGRs into programs for sustainable agriculture and crop improvement [2]. The characterization and maintenance of PGRs is therefore necessary to protect agrobiodiversity and to promote the development of climate-resilient crops and agroecosystems [2]. Core collections offer a useful concept for the characterization of diversity preserved in PGRs [3, 4, 5]. A core collection is smaller than a comprehensive collection but maximizes diversity, and can therefore be used as an efficient kick-off point to enhance genetic gains, including the use of genomics, phenomics, and molecular phenotyping [1, 6, 7, 8, 9]. Core collections have been developed for many legumes, including soybean [10], common bean [6, 11, 12], cowpea [13], pigeon pea [14, 15], groundnut [16], chickpea [17], lentil [18, 19, 20], and lablab bean [21]. Many grain legumes have been explored in detail, but overreliance on a few major legumes with poor climate resilience threatens sustainable food systems by accelerating resource depletion and environmental harm [22]. To enhance food and nutritional security, it is necessary to promote the cultivation and utilization of orphan legumes–such as neglected or underutilized lupin species– due to their ecological and nutritional potential. These crops can contribute significantly to sustainable agriculture and environmental resilience.

Lupins (*Lupinus* spp.) belong to the legume family (Fabaceae) and grow all around the world, having adapted to multiple climate zones [23]. Lupin seeds are rich in protein, oil, carbohydrates, and fiber, making them nutritious plant-based foods [1, 24, 25]. They are also valuable as feed for livestock and aquaculture, and material for soil enrichment [26, 27]. Lupin species have attracted recent interest from researchers, particularly in the field of genomics [28, 29, 30, 31]. The cultivated species are *Lupinus albus* (white lupin), *L. angustifolius* (blue lupin), *L. luteus* (yellow lupin), and *L. mutabilis* (pearl lupin). Among these crops, *L. albus* is particularly notable due to its 30–40% seed protein content [32] and adaptability to different climate and soil conditions [33]. It recently joined *L. angustifolius* and *L. luteus* in the EU Novel Food status Catalogue (European Commission, 2024). The additional proposed dietary benefits of lupin seeds include the prevention of diabetes and cardiovascular disease, and the maintenance of a healthy digestive system due to the high fiber content [24, 34].

Although *L. albus* is a promising legume crop, its genetic resources are poorly characterized. Currently, 5564 accessions (only 915 of which are available for distribution) are held in gene banks (https://www.genesys-pgr.org/a/overview/v293JdOPwZG, accessed on 21.08.2024), but few studies thus far have considered the diversity of these collections [35, 36, 37]. Even so, substantial phenotypic diversity has been documented for several traits, many of which vary among accessions from the same geographical region or country of origin, according to passport data [33, 37, 38, 39]. For example, among *L. albus* ecotypes from 17 countries, the pod wall to whole pod weight ratio was highest for Greek and Italian ecotypes, and lowest for Egyptian ecotypes [40]. Furthermore, 45 local Greek landraces showed variation in seed weight (lowest-to-highest ratio up to 12-fold) and size (smallest-to-largest ratio up to 3-fold) [39]. Among 35 spring-sown Spanish landraces, only three of 50 quantitative parameters (related to leaf morphology) showed a coefficient of variation (CV) ≥ 19%, and only four of 51 qualitative descriptors (related to flower color) showed variation between accessions [41]. Variations have also been observed in Ethiopian lines for seed minerals such as Ca (709–1284 μg/g), P (671.3–2490.2 μg/g), Fe (77.9–92.8 μg/g), and Mg (1739–2159 μg/g) [42, 43], and in protein content, with a range of 28.55–35.81% [44]. These findings highlight the substantial untapped diversity in *L. albus* germplasm, but systematic characterization and accessibility remains fragmented and limited in scope. To the best of our knowledge, two core collections of *L. albus* have been developed: 248 Spanish lines based on ecogeographical data [45], and 34 Ethiopian lines based on microsatellite markers [33]. Therefore, broader and more integrated collections are needed to support breeding and conservation programs more effectively.

To advance the core collection concept, a novel framework known as *Intelligent Collections* has been introduced for the conservation, management, and analysis of genetic resources [1] (Fig. 1A). These collections are dynamic (i.e., able to memorize, learn, improve, and evolve) and represent the entire genetic diversity of four key legume crops: common bean, lentil, lupin, and chickpea. *Intelligent Collections* are designed as a set of nested core collections with different sizes: a Reference-CORE (R-CORE) comprising thousands of single-seed descent (SSD) lines representing the total genetic resources of the species, a Training-CORE (T-CORE) comprising a subset of R-CORE lines, usually a few hundred, and a Hyper-CORE (H-CORE) featuring 40–50 genotypes sampled on an evolutionary transect [1]. As part of the INCREASE project [1], multi-omics approaches (including genomics, transcriptomics, and metabolomics) and molecular phenotyping are applied to these collections to characterize them in more detail. Using this approach, a large

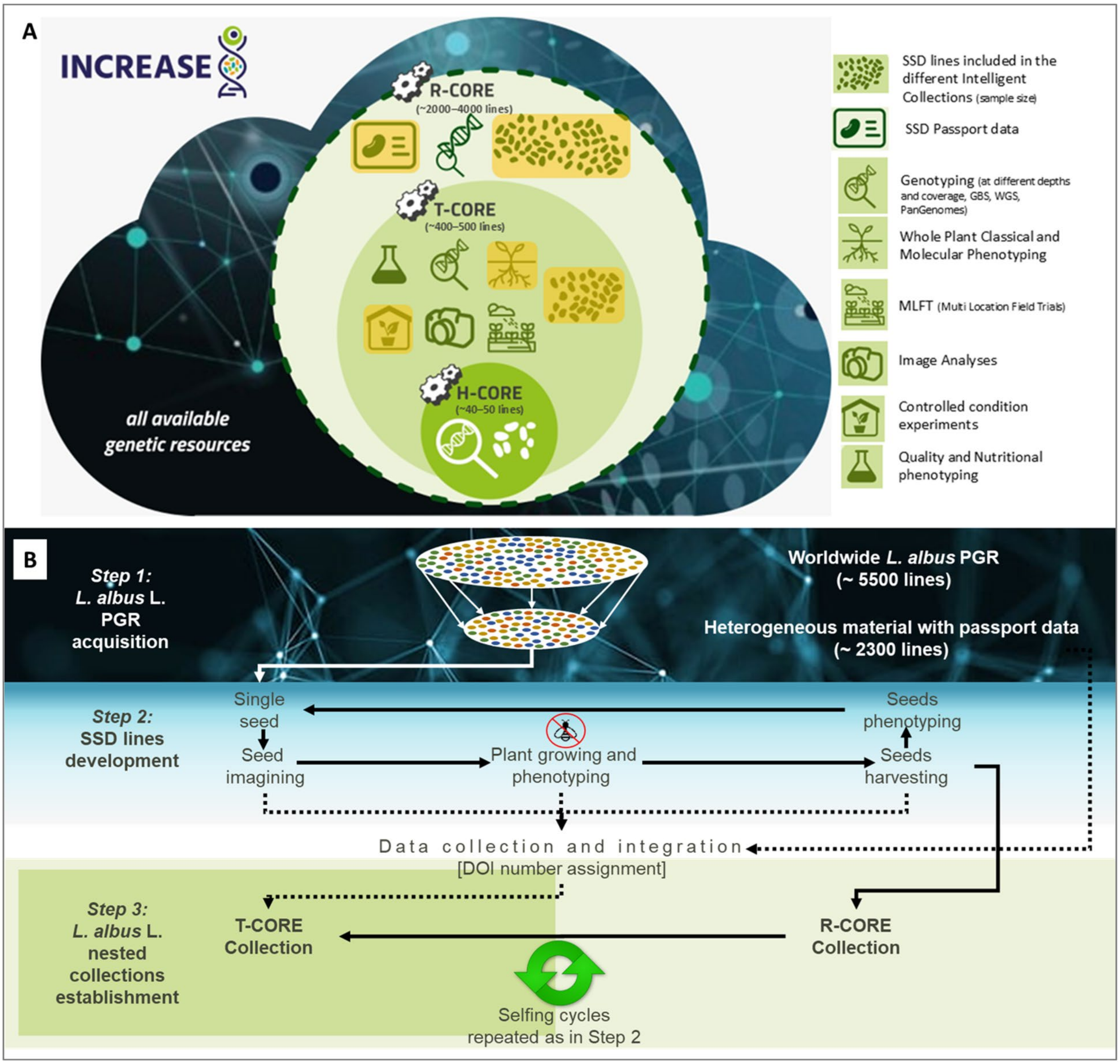

Fig. 1 Core collections and single-seed descent (SSD) lines. (A) Schematic representation of the INCREASE nested core collection (*Intelligent Collections*) comprising the Reference-CORE (R-CORE), Training-CORE (T-CORE) and Hyper-CORE (H-CORE) modified from Ref [1]. (B) Development of the *Lupinus albus* SSD lines and seed multiplication cycles. The number of accessions in global *L. albus* germplasm is based on Genesys PGR (https://www.genesys-pgr.org/a/overview/v293JdOPwZG, accessed on 21.08.2024)

(~10,000) collection of common bean accessions and a subsample of 500 accessions (Pv_core1) have been characterized using the genotyping-by-sequencing approach. A further subsample of 220 accessions has been genotyped by whole-genome sequencing [6, 46]. In parallel, a core set of 480 SSD chickpea genotypes has been evaluated for phenotypic traits [47].

In this study, we report the development and phenotypic characterization of the first nested core collections of *L. albus* according to the *Intelligent Collections* strategy [1] (Fig. 1B). These scalable, structured collections represent the full spectrum of *L. albus* diversity. Classical phenotyping across this nested set will lay the groundwork for the integration of genetic, genomic, and molecular data. Our aim is to support the sustainable utilization of *L. albus* genetic resources in an integrative and data-driven approach for the broader goal of enhancing the sustainable use of legume PGRs in the development of agricultural systems that are resilient, diverse, and focused on nutrition.

Methods

Plant material and SSD line development for the core collections

Of the 5564 *L. albus* accessions in gene banks, we obtained 2288 accessions from gene banks, INCREASE project partners: Leibniz-Institut für Pflanzengenetik und Kulturpflanzenforschung (IPK, Germany), Università Politecnica delle Marche (UNIVPM, Italy), ISEA Srl (currently, Società Produttori Sementi Spa, Italy), Consiglio per la Ricera in Agricoltura e l'Analisi dell'Economia Agraria (CREA, Italy), Instytut Hodowli i Aklimayzacji Roslin - Panstwowy Instytut Badawczy (IHAR-PIB, Poland), Banca de Resurse Genetice Vegetale „Mihai Cristea" Suceava (BRGV Suceava, Romania), Statiunea de Cercetare Dezvoltare Pentru Legumicultura Bacau (SCDL-BACAU, Romania), stakeholders: Jouffray-Drillaud (currently Cérience, France), Millennium Seed Bank Kew Garden (MSB Kew Garden, UK) and research institutes and gene banks: Australian Grains Genebank (AGG, Australia), Gene bank — Czech Republic (GRIN CZECH, Czech Republic), Centre for Genetic Resources the Netherlands (CGN, Netherlands), Centro de Investigaciones Científicas y Tecnológicas de Extremadura (CICYTEX, Spain), United States Department of Agriculture Agricultural Research Service (USDA ARS) (Additional File 1: Table S1) for our core collections. SSD line development and phenotypic characterization have already been described in detail [25]. Briefly, SSD lines were derived from one randomly chosen seed from each accession, and after one selfing cycle these 2288 SSD lines constituted the R-CORE [1, 25]. Phenotyping was conducted successively during the plant growth cycle. Also for each SSD line, unique Digital Object Identifier (DOIs) are being assigned for data integration, and is facilitated by International Treaty on Plant Genetic Resources for Food and Agriculture. We are repeating these selfing cycles for seed multiplication, expanding the SSD lines to conserve and share PGRs in the future.

Table 1 Qualitative traits used for the phenotypic characterization of *Lupinus albus* [25]

No.	Descriptors	Code	Descriptor state
1	Color of wing petals at full flowering on the main stem	F6	Blue, pink, violet, white, white/blue, white/pink, white/violet
2	Color of standard petal at full flowering on the main stem	F7	Blue, pink, violet, white, white/blue, white/pink, white/violet
3	Color of the keel edge at full flowering on the main stem	F8	Brownish, not visible, transparent, white, white/blue
4	Purple pigmentation of the main stem after two months of emergence	PP	Medium intensity, no pigment plus purple internodes, none, slightly, very intense
5	Pod shattering on the main stem	P4_m	Complete, moderate, none, slight
6	Pod shattering on the lateral branches	P4_l	Moderate, none, slight
7	Seed shape	S5	Cuboid, flattened oval, flattened cuboid, flattened spherical
8	Seed ornamentation	S6	Marbled, none, other
9	Seed ornamentation color	S7	Black, dark brown, light brown, none, other
10	Density of seed ornamentation	S8	Dense, medium, none, very dense
11	Growth habit	GH	Determinate, indeterminate, semi-determinate

The R-CORE SSD lines were grown in the greenhouse at the Institute of Plant Genetics, Polish Academy of Sciences (IPG-PAN), Poznan, Poland (52°26′50.5′′–16°54′13.1′′) with a 16-h photoperiod and 60–65% relative humidity. T-CORE plants (300 accessions selected from the R-CORE) were grown and phenotyped in three experiments at two locations: first, the IPG-PAN greenhouse (coordinates as above) from October 2020 to May 2021 (experiment ID P1C20_la2) and from March to August 2021 (experiment ID P1C21_la1), and second at the greenhouse of the Agricultural Institute of Slovenia (KIS), Ljubljana, Slovenia (46°8′42′′–14°33′28.79′′) from March to August 2021 (experiment ID S1C21_la). In all three T-CORE experiments, we applied a 16-h photoperiod and an average 65% relative humidity. For both core collections, phenotyping was carried out during each growing cycle, and images of qualified seeds from each accession were captured at the harvesting stage for processing [25]. The seeds were then stored.

Phenotyping of agro-morphological traits and data analysis

We assessed 35 agro-morphological traits for phenotypic characterization: 11 qualitative (Table 1) and 24 quantitative (Table 2) [25]. Phenotypic relationships among accessions were determined by the multivariate analysis of these traits using R v4.3.1. First, qualitative traits in the R-CORE collection were characterized to understand the overall variation in the species. Principal component analysis (PCA) using binary data was then applied using the *PCA ()* function in the "FactoMineR" package [48] to retain the most meaningful components. The clustering tendency in the dataset was tested with Hopkins statistics *H*, using the *hopkins ()* function in the "clustertend" package [49]. We performed hierarchical clustering on the retained principal components (HCPC) using the *HCPC ()* function in the "FactoMineR" package. Graphical outputs were visualized using the function *fviz_cluster ()* in the "factoextra" package [50].

To characterize the phenotypic diversity among the 24 quantitative traits, descriptive statistics (minimum, maximum, mean, standard deviation, CV, and frequency

Table 2 Quantitative traits used for the phenotypic characterization of *Lupinus albus* [25]

Phenotypic trait /phenological stage	No.	Descriptors	Code	Unit
Emergence	1	Days to emergence	Emer	Days
Flowering	2	Days to flowering (length of vegetative phase)	F1_m	Days
	3	Days to first petal occurrence on the main stem	F2_m	Days
	4	Days to first flower occurrence on the main stem	F3_m	Days
	5	Days to the full flowering on the main stem	F4_m	Days
	6	Days to the first flower withering on the main stem	F5_m	Days
	7	Length of the generative phase on the main stem I	Gen	Days
Pod development and physiological maturity	8	Days to the first pod set on the main stem	P1_m	Days
	9	Days to the first pod set on the lateral branches	P1_l	Days
	10	Days to the first pod color changes on the main stem	P2_m	Days
	11	Days to the first pod color changes on the lateral branches	P2_l	Days
	12	Length of the pod and seed growth phase on the main stem I	Mat1_m	Days
	13	Length of the pod and seed growth phase on the main stem II	Mat2_m	Days
	14	Length of the pod and seed growth phase on the lateral branches I	Mat1_l	Days
	15	Length of the pod and seed growth phase on the lateral branches II	Mat2_l	Days
Pods and seeds production	16	Number of pods per plant on the main stem	P3_m	Numbers
	17	Number of pods per plant on the lateral branches	P3_l	Numbers
	18	Number of qualified seeds on the main stem	S1_m	Numbers
	19	Number of qualified seeds on the lateral branches	S1_l	Numbers
	20	Total number of qualified seeds	S2	Numbers
	21	Total mass of the qualified seeds	S3	g
	22	1000 seed weight	S4	g
Plant height	23	Plant height of the main stem	H1	cm
	24	Total plant height	H2	cm

distribution) were computed and the normality was tested using Kolmogorov–Smirnov and Shapiro–Wilk normality tests. The quantitative traits were assessed by PCA using the "FactoMineR" package and one-way analysis of variance (ANOVA) followed by Tukey's multiple comparisons test ($P \leq 0.01$) to explore differences among accessions representing different phenotypic groups, geographical regions, and biological statuses (cultivar, landrace, wild, breeding/research material). Pearson's or Spearman's correlation tests (as appropriate) were applied to find significant relationships between qualitative and quantitative traits using the "corrplot" and "GGally" packages. We also used packages such as "ggplot2" (graphic presentation), "sf", "rworldxtra" and "rnaturalearth" (for mapping and geo-referencing).

To estimate genetic (V_g) and residual (V_r) variance for each trait, we applied a restricted maximum likelihood (REML) linear mixed model (LMM) in the package "lme4" [51] as follows: $Yi = \mu + Ai + \varepsilon i$, where Yi is the response variable of accession i, μ is the trait mean, Ai is the accession's total genetic value, and εi is the residual value. For comparison across traits, phenotypic groups, geographical regions and biological statuses, we calculated the coefficient of total genetic variation (CV_g) using the equation $CV_g = \sqrt{V_g}/\mu$, where V_g is the total genetic variance and μ is the mean of the trait [52].

Development and characterization of the T-CORE as a subset of the R-CORE

We selected 300 R-CORE accessions representing maximum diversity to establish the T-CORE, based on R-CORE passport data (biological status, geographical origin, and collection site) and phenotypic data (Tables 1 and 2). Two sampling strategies were used. First, random-stratified sampling (logarithm of frequency of accessions by geographical region, country, and biological status) was used to determine the number of accessions from each group to populate the T-CORE. Second, accessions from these groups were sampled based on phenotypic data to capture the maximum phenotypic diversity of the R-CORE [1]. The phenotypic data were standardized to eliminate scale differences, and different combinations of variables were used to establish and assess the T-CORE according to qualitative traits, quantitative traits, and a combination of quantitative and qualitative traits. Finally, to ensure maximum diversity in the T-CORE, certain accessions were manually added or excluded. For example, a greater number of accessions from Greece and Middle Eastern countries were included because these regions are potential centers of lupin domestication. Conversely, the number of landraces was reduced because most originated from Spain and were overrepresented in the initial T-CORE.

To ensure the T-CORE collection was representative of the R-CORE collection, we applied 'EvaluateCore' v0.1.2 [53] to determine the frequency distributions for geographical regions, countries or origin, and biological status, as well as qualitative traits, using a Chi-squared (χ^2) test for homogeneity. Coverage of the T-CORE qualitative traits was estimated using the equation for class coverage: $CC\,\% = \left(\frac{1}{n}\sum_{n=1}^{n}\frac{\text{Ati}}{\text{Ari}}\right) * 100$ [54], where At_i and Ar_i are the number of categories for the i^{th} trait observed in the T-CORE and R-CORE collections, respectively, and n is the number of qualitative traits. We applied the Shannon–Weaver diversity index [55] to measure phenotypic diversity for each qualitative trait in both collections using the formula $\text{H}' = -\sum_{i}^{n} pi\text{log}pi$, where p_i is the frequency of accessions from the i^{th} class and n is the number of classes for the designated trait. Furthermore, we computed Pearson's coefficient (r) between quantitative traits and Cramer's V coefficient (v) between qualitative traits in the R-CORE and TCORE using packages 'corrplot' [56] and 'vcd' [57], respectively, to assess the extent of trait associations captured in the T-CORE collection.

Results

Development and characterization of the *L. albus* R-CORE collection

We used 2288 accessions to produce SSD lines after one selfing cycle for the *L. albus* R-CORE collection, of which 1178 SSD lines have been assigned with unique DOIs so far. According to passport data, these accessions originated from 43 countries distributed over nine geographical regions (Fig. 2A; Table 3). Most of the accessions (76.4%) were from Europe, mainly Southern Europe (1336 accessions, 58.4%), followed by Western Europe (209, 9.1%), and Eastern Europe (203, 8.9%). Outside Europe, most accessions originated from Northern Africa (208, 9.1%), followed by the Middle East (47, 2.1%), America (46, 2%), and Eastern Africa (40, 1.7%). Australia and New Zealand (11, 0.5%) and Southern Asia (1, 0.04%) were the least representative regions. Another 187 accessions (8.2%) were described as "unknown origin". In the Southern Europe category, most accessions originated from Spain (615, 26.8%) and Portugal (439, 19.2%).

Passport data (Fig. 2B; Table 4) revealed that the majority of the R-CORE collection was made up of landraces (1173, 51.3%), cultivars (392, 17.1%) and wild accessions (233, 10.2%), as well as breeding/research materials (176, 7.7%). However, the biological status of 314 (13.7%) accessions was unknown. The majority of breeding/research materials and cultivars originated from Europe (Eastern, Western, and Southern). The landraces were mainly represented by Southern Europe and Northern Africa, whereas the wild accessions mainly originated from Southern Europe and the Middle East (Fig. 2B, Additional File 2: Figure S1).

Characterization of the R-CORE collection based on qualitative traits

The R-CORE collection showed considerable diversity for most of the qualitative traits (Fig. 3, Additional File 1: Tables S2 and S3). The flower color traits at the full flowering stage, namely the color of the wing petal (F6) and standard petal (F7), feature seven descriptors. For the wing petals, the most common colors were white/blue and white/violet (1451 and 383 accessions, respectively), whereas white and white/blue were the most common for the standard petals (1103 and 970 accessions, respectively). Moreover, for most accessions, the keel edge (F8) was either invisible (1594) or transparent (662). In terms of stem pigmentation, most of the accessions (1450) featured purple internodes only, whereas in others the pigmentation was either absent (439), or slightly (210), medium (140), or very intense (49). In terms of seed shape (S5), flattened spherical (1268) and flattened cuboid (846) were the most common. All accessions formed white seeds, in most cases non-ornamented (2237). Ornamented seeds were identified in only 51 accessions, 50 of which were of the marbled type. We observed growth habit (GH) using three categories: determinate, with inflorescence on the main stem (58 accessions); semi-determinate, with primary lateral branches along the main stem (445); and indeterminate, with further branching on secondary lateral branches (1785). Statistical analysis for all qualitative traits revealed maximum diversity for the color of wing petals (F6), the purple pigmentation on the main stem (PP), and the color of the standard petals (F7), with H′ values of 1.15, 1.08 and 1.01, respectively (Table 5). Minimum diversity was observed for pod shattering on the main stem (P4_m) and on lateral branches (P4_l), with H′ values of 0.07 and 0.02, respectively.

Characterization of the R-CORE collection based on quantitative traits

The R-CORE collection also showed considerable variability for most of the quantitative traits, and revealed a non-normal distribution in all cases (Fig. 4). Descriptive statistics are presented in Table 6. The 1000 seed weight showed the widest range (S4; range 15–680 g; mean ± SD = 188.3 ± 70.55 g). This was followed by the length of the pod and seed growth phase on the main stem II (Mat2_m; 3.75–156 days; 115.46 ± 39.77days), the length of the pod and seed growth phase on the main stem I (Mat1_m; 13.18–136 days; 107.94 ± 31.51 days), and total plant height (H2; 17–134 cm; 70.28 ± 18.35 cm). The smallest range was observed for the number of pods per plant on the main stem (P3_m; 0–11; 1.91 ± 1.76) and the number of pods per plant on the lateral branches

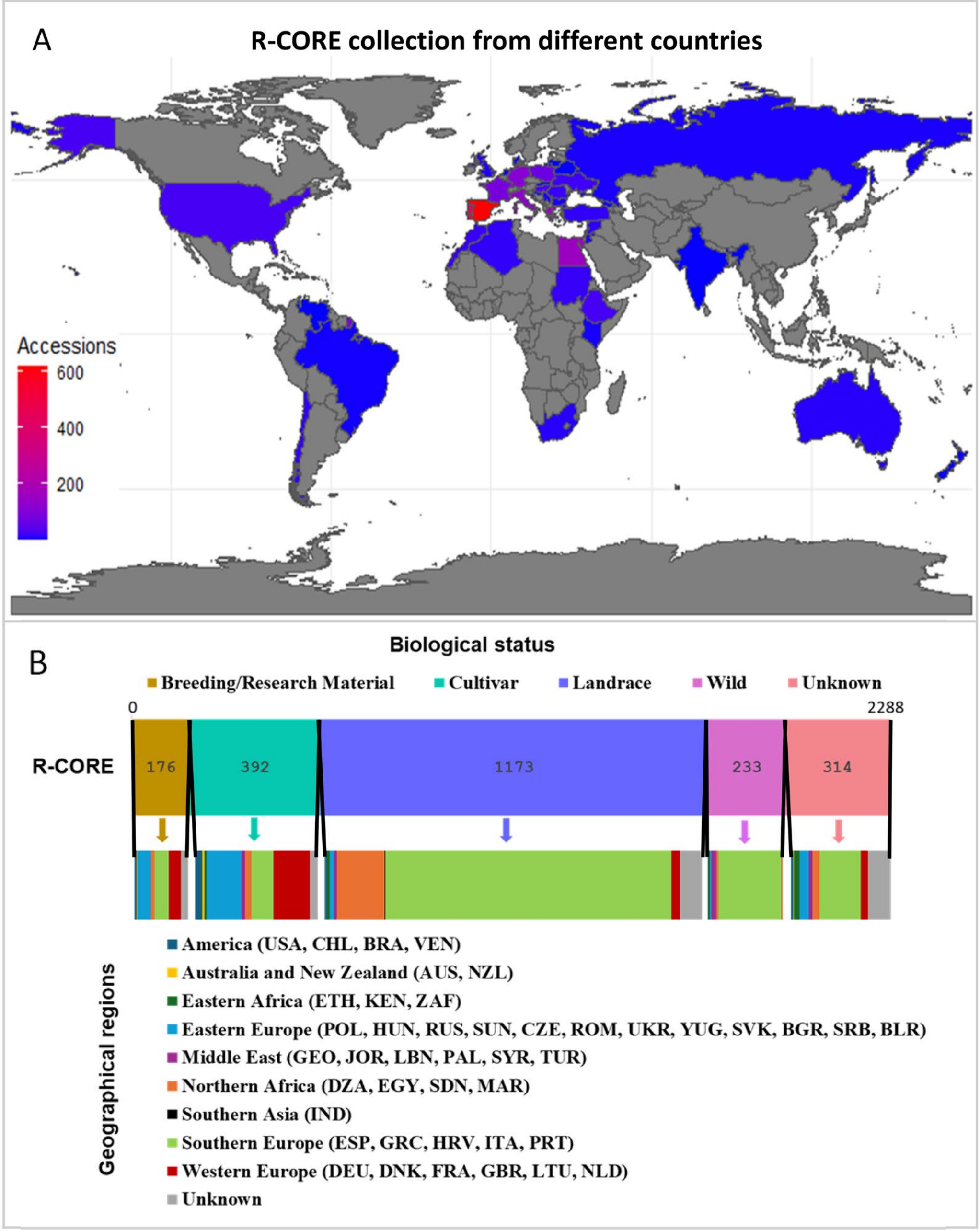

Fig. 2 Characterization of *Lupinus albus* accessions in the R-CORE collection based on (A) country of origin and (B) biological status, subdivided by geographical region

(P3_l; 0–16; 2.88 ± 3.01). The highest CVs were observed for the number of seeds on lateral branches (S1_l; CV = 150.24%), the total mass of qualified seeds (S3; 120.85%), and the number of seeds on the main stem (S1_m; 116.25%). The lowest CVs were observed for the number of days to the first pod color change on the main stem (P2_m; 9.18%) and days to the first pod set on the main stem (P1_m; 11.14%).

Development and characterization of the *L. albus* T-CORE collection

A set of 300 accessions from the *L. albus* R-CORE collection was used to build the T-CORE collection (Additional File 1: Table S1). The genetic diversity of the 300 T-CORE accessions was comparable to that of the R-CORE accessions as reflected by the relative distribution of the geo-regions/country of origin and biological status of these accessions (Tables 3 and 4).

Table 3 Number (N) and percentage of accessions from different geographical regions (subdivided by country$^{\#}$) in the *L. albus* R-CORE and T-CORE collections

Geographical region$^{\#}$	**R-CORE**		**T-CORE**		**χ^2**	**p-value**
	(N = 2288)		**(N = 300)**			
	N	**%**	**N**	**%**		
Southern Europe	1336	58.4	173	57.7	0	ns
Spain (ESP)	615	26.8	96	32	4	0.049 *
Greece (GRC)	152	6.6	24	8	0.8	ns
Croatia (HRV)	2	0.1	0	0	0.3	ns
Italy (ITA)	128	5.6	40	13.3	28.4	<0.001**
Portugal (PRT)	439	19.2	13	4.3	40.5	<0.001**
Western Europe	209	9.1	33	11	0.5	ns
Germany (DEU)	110	4.8	24	7.7	5	<0.05*
Denmark (DNK)	1	0	0	0	0.1	ns
France (FRA)	86	3.8	6	1.7	3.7	ns
United Kingdom (GBR)	7	0.3	1	0.3	0	ns
Lithuania (LTU)	1	0	1	0.3	2.9	ns
Netherlands (NLD)	4	0.2	1	0.3	0.4	ns
Northern Africa	208	9.1	21	7	1.4	ns
Algeria (DZA)	14	0.6	4	1.3	2.1	ns
Egypt (EGY)	159	6.9	12	4	3.7	ns
Sudan (SDN)	17	0.7	3	1	0.2	ns
Morocco (MAR)	18	0.8	2	0.7	0.1	ns
Eastern Europe	203	8.9	25	8.3	0.1	ns
Poland (POL)	67	2.9	10	3.3	0.2	ns
Hungary (HUN)	28	1.2	5	1.7	0.4	ns
Russia (RUS)	5	0.2	1	0.3	0.2	ns
USSR (SUN)	34	1.5	4	1.3	0	ns
Czech Republic (CZE)	9	0.4	1	0.3	0	ns
Romania (ROM)	18	0.8	0	0	2.5	ns
Ukraine (UKR)	22	1	2	0.7	0.3	ns
Serbia and Montenegro (YUG)	9	0.4	2	0.7	0.5	ns
Slovakia (SVK)	1	0	0	0	0.1	ns
Bulgaria (BGR)	5	0.2	0	0	0.7	ns
Serbia (SRB)	4	0.2	0	0	0.5	ns
Belarus (BLR)	1	0	0	0	0.1	ns
Middle East	47	2.1	17	5.7	14.4	<0.001**
Georgia (GEO)	2	0.1	1	0.3	13.9	ns
Jordan (JOR)	5	0.2	2	0.7	2	ns
Lebanon (LBN)	2	0.1	0	0	0.3	ns
Palestine (PAL)	10	0.4	6	2	10.6	<0.001**
Syria (SYR)	14	0.6	2	0.7	0	ns
Turkey (TUR)	14	0.6	6	2	6.7	<0.05*
America	46	2	3	1	1.5	ns
Chile (CHL)	15	0.7	2	0.7	0	ns
Brazil (BRA)	3	0.1	0	0	0.4	ns
Venezuela (VEN)	1	0	0	0	0.1	ns
Eastern Africa	40	1.7	7	2.3	0.4	ns
Ethiopia (ETH)	27	1.2	5	1.7	0.5	ns
Kenya (KEN)	4	0.2	0	0	0.5	ns
South Africa (ZAF)	9	0.4	2	0.7	0.5	ns
Australia and New Zealand	11	0.5	1	0.3	0.1	ns
Australia (AUS)	9	0.4	1	0.3	0	ns
New Zealand (NZL)	2	0.1	0	0	0.3	ns

Table 3 (continued)

Geographical region[#]	**R-CORE**		**T-CORE**		χ^2	**p-value**
	(N = 2288)		**(N = 300)**			
	N	**%**	**N**	**%**		
Southern Asia	1	0.04	0	0	0.1	ns
India (IND)						
Unknown	187	8.2	20	6.7	0.9	ns

Chi squared analysis (χ^2) was used to determine the goodness-of-fit of observed values with theoretical expected values for each geographical region/country. *p*-values indicate statistical significance at $p<0.05$.

ns nonsignificant

[#] The names of countries are as provided in the passport data obtained from the gene banks

Table 4 Number, percentage, and χ^2 value of accessions representing different biological statuses in the *L. albus* R-CORE and T-CORE collections, based on passport data

Biological status	***R*-CORE**		**T-CORE**		χ^2	***p*-value**
	N	**%**	***N***	**%**		
Cultivar	392	17.1	52	17.3	0.0	ns
Landrace	1173	51.3	89	29.7	51.0	< 0.001**
Breeding/research material	176	7.7	42	14.0	15.2	< 0.001**
Wild accession	233	10.2	75	25.0	58.6	< 0.001**
Unknown	314	13.7	42	14.0	0.0	ns

Chi squared analysis (χ^2) was used to determine the goodness-of-fit of observed values with theoretical expected values for each biological status. *p*-values indicate statistical significance at $p<0.05$.

ns nonsignificant

The T-CORE collection comprised 89 landraces (29.7%), 75 wild accessions (25%), 52 cultivars (17.3%) and 42 breeding/research materials (14%) from 30 countries representing eight geographical regions (Table 4; Fig. 5A, Additional File 1: Table S1). Another 42 accessions (14%) lacked an assigned biological status and the geographical origin of 20 (6.7%) was unknown. Five accessions (1.7%) lacked both designations. Like the R-CORE, most T-CORE accessions (231, 77%) originated from Europe, mainly Southern Europe (173, 57.7%), followed by Western Europe (33, 11%), and Eastern Europe (25, 8.3%). The major contributors outside Europe were Northern Africa (21, 7%), the Middle East (17, 5.7%), and East Africa (7, 2.3%). America (3, 1%) and Australia and New Zealand (1, 0.3%) were minor contributors. Among the accessions from Southern Europe, the majority originated from Spain (32%) and Italy (13.3%). The nonsignificant χ^2 value for all geographical regions and countries indicated that all regions were optimally represented in the T-CORE collection (Table 3). Most of the qualitative traits were homogenously distributed in the R-CORE and T-CORE collections, with a CC of 66.66–100% and mean of 82.16% for all traits (Table 5). For the quantitative traits, wide variability was again observed in T-CORE similar to that in the R-CORE (Table 6).

Correlation analysis in the T-CORE and R-CORE collections

The correlation patterns for most traits were similar in the R-CORE and T-CORE (Fig. 5). For the quantitative traits (Fig. 5C, Additional File 1: Table S4), we observed strong positive correlations ($r>0.70$) among the flowering-related traits (F1_m-F5_m) and relatively strong associations with traits related to pod development ($r>0.40$) in the T-CORE and R-CORE. For example, days to first flower occurrence on the main stem (F3_m) showed strong positive correlations with days to flowering (F1_m; $r=0.83$ in T-CORE, $r=0.91$ in R-CORE) and days to first petal occurrence on the main stem (F2_m; $r=0.93$ in T-CORE, $r=0.98$ in R-CORE) among the flowering traits, and with days to first pod set on the main stem (P1_m; $r=0.81$ in T-CORE, $r=0.83$ in R-CORE) and lateral branches (P1_l; $r=0.71$ in T-CORE, $r=0.63$ in R-CORE) among traits related to pod development. We also observed significant positive correlations among traits characterizing the physiological maturity of the main stem and lateral branches, but these traits showed significant negative correlations with all traits related to flowering. In addition, in both collections, small negative correlations ($r > -0.21$) were observed between flowering-related traits and traits related to the production of pods and seeds. The number of pods and qualified seeds on lateral branches showed a strong positive correlation with the total number of qualified seeds. For the qualitative traits, the strongest positive correlations were observed among traits related to seed ornamentation (S6, S7, and S8; $r>0.9$ in both collections) and between the color of wing petals and standard petals (F6, F7; $r=0.64$ in the T-CORE and $r=0.71$ in the R-CORE) (Fig. 5B, Additional File 1: Table S5). The growth habit (GH) showed a moderate positive correlation with pod shattering on

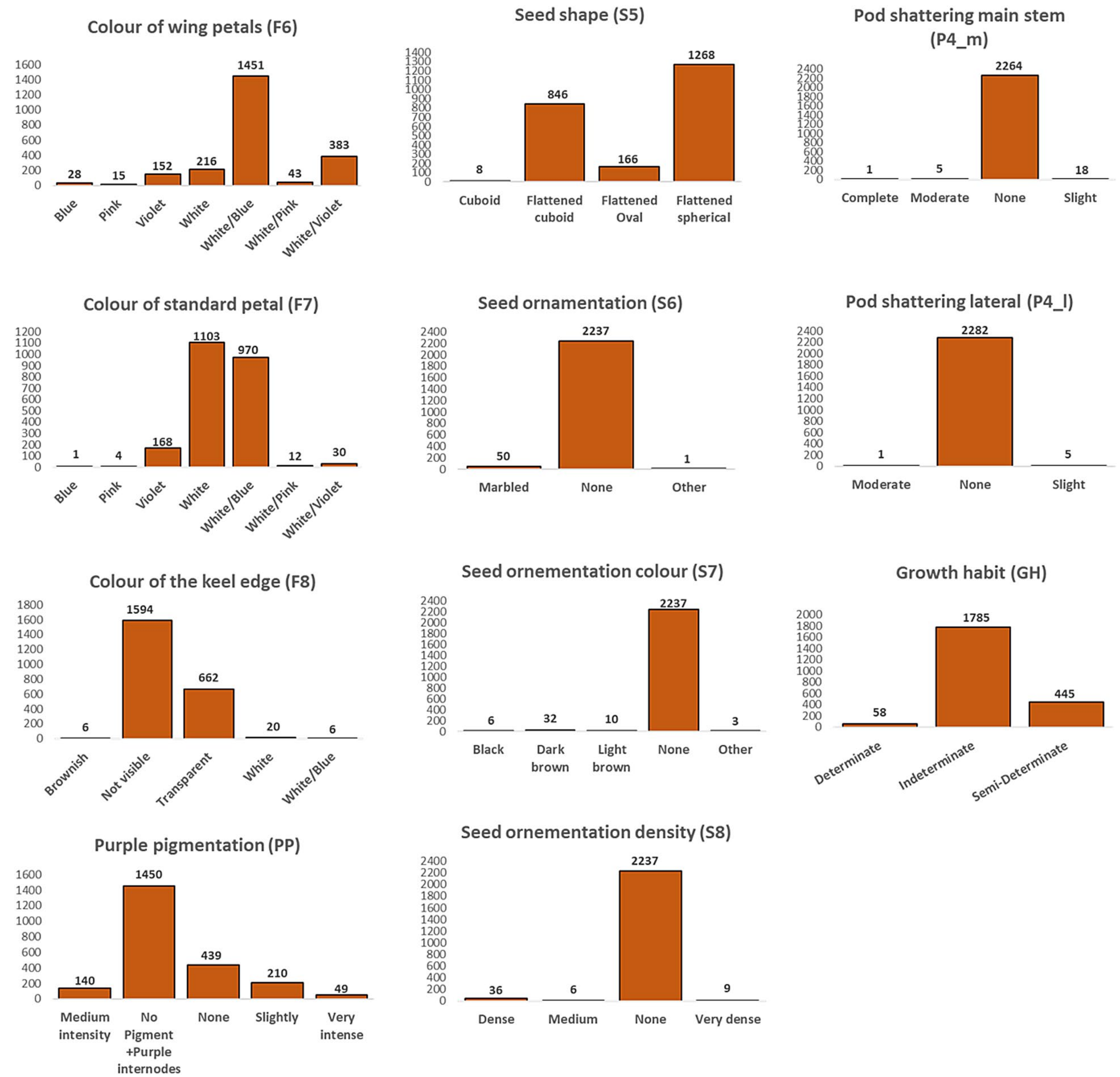

Fig. 3 Frequency distribution of qualitative traits in the *L. albus* R-CORE collection

the main stem (P4_m) and lateral branches (P4_l) in both collections ($r > 0.30$). Furthermore, we observed a small but significant positive correlation ($r < 0.20$) among the traits related to petal color (F6, F7), seed ornamentation (S6), seed ornamentation color (S7), and the density of ornamentation (S8) in both collections.

Phenotypic variability in the T-CORE collection

The results described above confirmed the wide variability of quantitative traits in the T-CORE collection, and a Shapiro–Wilk normality test showed that, except for plant height (H2; $p = 0.198$) and the number of qualified seeds on the main stem (S1_m; $p = 0.096$), all traits displayed non-normal distribution (Fig. 6). For example, days to flowering (F1_m) varied from 32 to 52 days, days to full flowering on the main stem (F4_m) ranged from 50 to 74 days, and days to first pod set ranged from 51 to 75 days on the main stem (P1_m) and from 61 to 86 days on the lateral branches (P1_l). For the traits related to pods and seeds, the number of pods ranged from 0 to 10 on the main stem (P3_m) and up to 15 on the lateral branches (P3_l), the number of qualified seeds ranged from 0 to 34 on the main stem (S1_m) and up to 51 on the lateral branches (S1_l), and the total number of qualified seeds ranged from 20 to 73 (Table 6). Interestingly, the wild accessions INLUP_00272 from Ukraine, INLUP_00131, INLUP_00135 and INLUP_00130 from Spain, INLUP_00233 from Turkey, INLUP_00136 from

Table 5 Patterns of variation and χ^2 values for the frequency distributions of qualitative descriptors in the *L. albus* R-CORE and T-CORE collections

	Trait	R-CORE			T-CORE			CC (%)	χ^2	p-value
		H′	H′ max	J′	H′	H′ max	J′			
Color of wing petals at full flowering on the main stem	F6	1.15	1.95	0.59	0.84	1.79	0.47	100	24.15	0.001 **
Color of standard petals at full flowering on the main stem	F7	1.01	1.95	0.52	0.82	1.39	0.59	57.14	16.6	0.02 *
Color of the keel edge at full flowering on the main stem	F8	0.68	1.61	0.42	0.25	1.39	0.18	100	224.16	0.000 ***
Purple pigmentation of the main stem after two months of emergence	PP	1.08	1.61	0.67	1.37	1.61	0.85	100	110.03	0.000 ***
Pod shattering on the main stem	P4_m	0.07	1.39	0.05	0.20	1.10	0.18	75	0.48	0.92 ns
Pod shattering on the lateral branches	P4_l	0.02	1.10	0.02	0.09	1.10	0.08	66.66	0.48	0.91 ns
Seed shape	S5	0.91	1.39	0.65	0.81	1.39	0.59	75	0.73	0.91 ns
Seed ornamentation	S6	0.11	1.10	0.10	0.11	0.69	0.16	75	0.73	0.91 ns
Seed ornamentation color	S7	0.13	1.61	0.08	0.13	1.39	0.10	80	3.63	0.41 ns
Density of seed ornamentation	S8	0.12	1.39	0.09	0.12	1.10	0.11	75	0.73	0.91 ns
Growth habit	GH	0.61	1.10	0.55	0.07	0.69	0.10	100	10.7	0.005 **
	Overall	0.53	1.47	0.34	0.44	1.24	0.31	82.16		

H′, H′$_{max}$ and J′ values indicate the Shannon–Weaver diversity index, maximum diversity and H′-based evenness, respectively (*p*-values indicate statistical significance at $p < 0.05$)

CC (%) coverage class, *ns* nonsignificant

Portugal, and one landrace (INLUP_00201 from Poland) produced the highest number of seeds (>62; Additional File 1: Table S2).

The Shannon diversity index for qualitative traits was similar in the R-CORE and T-CORE collections (Table 5), although the mean value was slightly lower in the T-CORE collection (H′ = 0.44, compared to 0.53 for the R-CORE), indicating that overall diversity was maintained in the T-CORE. When combined with the frequency distribution of qualitative traits (Additional File 1: Tables S2 and S3), these data show that the wing petal color (F6; H′ = 0.84) was white/blue in 225 accessions, whereas the standard petal (F7) was white/blue in 178 accessions and white in 110 accessions (H′ = 0.82). The color of the keel edge (F8; H′ = 0.25) was transparent for 284 accessions. Moreover, the purple pigmentation on the main stem (PP; H′ =1.37) was mainly characterized by slight pigmentation (108 accessions) and pigmentation at inter-nodes (103 accessions). In most T-CORE accessions, pod shattering was totally absent on the main stem (P4_m; 287 accessions) and lateral branches (P4_l; 295 accessions). In addition, the growth habit (GH; H′ = 0.07) was indeterminate in 296 accessions and semi-determinate in four. A determinate growth habit was not observed in any T-CORE accessions. Most accessions produced non-ornamented seeds (293 accessions) with flattened spherical (184 accessions) or flattened cuboid (103 accessions) shapes (Additional File 1: Tables S2 and S3).

Genetic diversity and phenotypic structure of the R-CORE collection

PCA applied to the 11 qualitative traits revealed that the two first components explained 19.3% (12.3% and 7%, respectively) of the total variance and divided the R-CORE accessions into two groups: one with non-ornamented seeds (group A, 2237 accessions), and a smaller one with ornamented seeds (group B, 51 accessions) (Fig. 7A, Additional File 1: Table S6). Group A formed three subgroups designated A1, A2 and A3 (Fig. 7B, Additional File 1: Table S7). Group A1 (2032 accessions) mainly featured white or white/blue flowers, a transparent keel, purple pigmentation on the stem, and indeterminate growth. It was made up of 176 wild accessions, 1047 landraces, 367 cultivars, 166 breeding/research materials and 276 accessions with unknown biological status, originating from all nine geographical regions (Fig. 8A). Group A2 (180 accessions) mainly featured violet flowers, an invisible keel, no purple pigmentation on the stem, non-shattering pods, and semi-determinate growth. It was made up of 26 wild accessions, 114 landraces, 18 cultivars, six breeding/research materials and 16 accessions with unknown biological status, again originating from all nine geographical regions. The smallest subgroup A3 (25 accessions) was mainly distinguished by shattering pods and white/blue flowers (Fig. 7B, Additional File 1: Table S7). It was made up of seven wild accessions, eight landraces, four cultivars, one breeding/research material, and five accessions with unknown biological status, and originated from Southern Europe (Italy and Spain), the Middle East (Palestine), Eastern Europe (Poland), and Northern Africa (Egypt) (Fig. 8A, B). Group B (accessions with ornamented seeds) was mainly represented by wild accessions (24), with only four landraces, three cultivars and three breeding/research materials. There were 17 accessions of unknown biological status. The accessions from group B originated from Europe, mostly (33 accessions) from Southern Europe– mainly Greece (Fig. 8B, Additional File 1: Table S8).

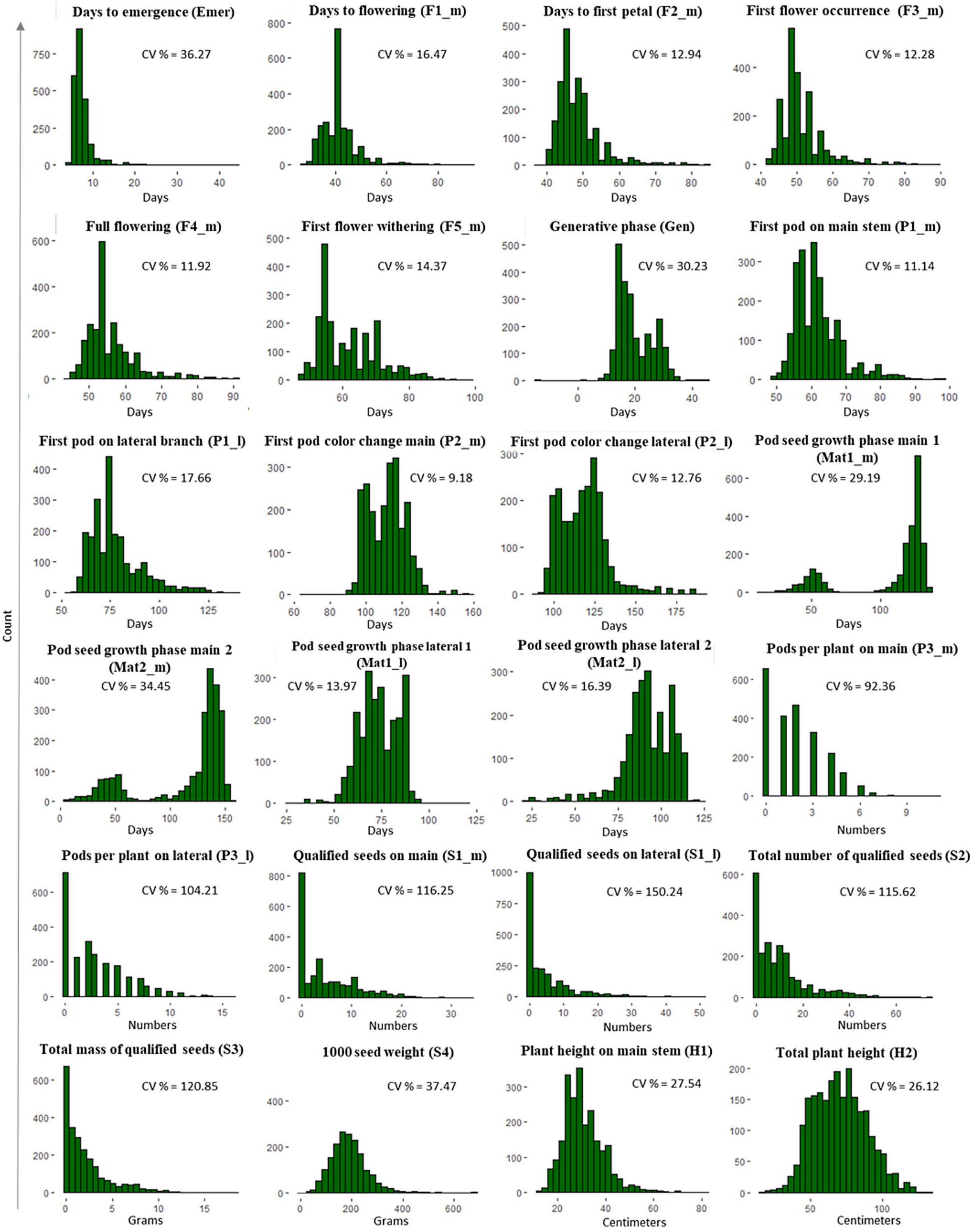

Fig. 4 Frequency distribution of quantitative traits in the *L. albus* R-CORE collection. *CV* coefficient of variation

Table 6 Statistical analysis of the quantitative traits in the R-CORE and T-CORE collections

	***R*-CORE**								**T-CORE**						
Quantitative trait	Code	min	max	mean	SD	CV (%)	median	IQR	min	max	mean	SD	CV (%)	median	IQR
Days to emergence	Emer	4	44	7.57	2.75	36.27	7	2	5	8	6.72	0.67	9.91	7	1
Days to flowering (length of vegetative phase)	F1_m	28	94	41.59	6.85	16.47	41	7	32	52	41.35	3.65	8.83	41	6
Days to first petal occurrence on the main stem	F2_m	37	84	48.66	6.3	12.94	47	6	40	63	49.54	3.7	7.46	49.5	5
Days to first flower occurrence on the main stem	F3_m	42	89	51.63	6.34	12.28	50	5	44	68	52.33	3.66	7	52	5
Days to the full flowering on the main stem	F4_m	44	91	55.66	6.63	11.92	53	6	50	74	57.06	4.11	7.21	57	7
Days to the first flower withering on the main stem	F5_m	47	98	61.61	8.85	14.37	60	13	56	82	67.83	4.56	6.73	67	4
Length of the generative phase on the main stem I	Gen	1.31	43.99	20.03	6.05	30.23	19	11	18	35	26.48	2.92	11.04	26	3
Days to the first pod set on the main stem	P1_m	49	98	61.93	6.9	11.14	60.51	8	51	75	61.43	4.5	7.33	60	5
Days to the first pod set on the lateral branches	P1_l	54	140	77.51	13.69	17.66	74	15.02	61	86	75.68	4.77	6.3	76	6
Days to the first pod color changes on the main stem	P2_m	64	158	111.65	10.25	9.18	112	15.58	105	130	119.91	4.16	3.47	119	6
Days to the first pod color changes on the lateral branches	P2_l	87	189.91	118.08	15.06	12.76	118	19.63	107	138	122.44	4.37	3.57	123	7
Length of the pod and seed growth phase on the main stem I	Mat1_m	13.18	136	107.94	31.51	29.19	123	17	47.26	134	122.67	9.33	7.61	125	5
Length of the pod and seed growth phase on the main stem II	Mat2_m	3.75	156	115.46	39.77	34.45	135	30.41	45.31	149	133.28	10.33	7.75	134	6
Length of the pod and seed growth phase on the lateral branches I	Mat1_l	27	121	73.35	10.25	13.97	73	15.58	55	80	65.09	4.16	6.39	66	6
Length of the pod and seed growth phase on the lateral branches II	Mat2_l	20.09	123	91.92	15.06	16.39	92	19.63	72	103	87.56	4.37	4.99	87	7
Number of pods per plant on the main stem	P3_m	0	11	1.91	1.76	92.39	2	3	0	10	3.84	1.55	40.48	4	2
Number of pods per plant on the lateral branches	P3_l	0	16	2.88	3.01	104.21	2	5	0	15	7.53	2.65	35.2	7	3
Number of qualified seeds on the main stem	S1_m	0	34	5.13	5.97	116.25	3	8	0	34	14.45	6.21	42.97	14.5	9
Number of qualified seeds on the lateral branches	S1_l	0	51	5.3	7.97	150.24	2	7	0	51	20.37	9.68	47.54	19.5	13
Total number of qualified seeds	S2	0	74	10.43	12.06	115.62	7	13	20	73	34.82	10.91	31.33	33	15
Total mass of the qualified seeds [g]	S3	0	18.37	2	2.42	120.85	1.2	2.7	2.56	18.37	6.85	2.27	33.11	6.64	2.78
1000 seed weight [g]	S4	15	680	188.3	70.55	37.47	182.37	83.11	108.75	409.57	199.56	42.4	21.25	194.14	47.41
Plant height of the main stem [cm]	H1	11	81	30.66	8.45	27.54	29.5	10.3	16	66	31.64	6.78	21.43	31	8
Total plant height [cm]	H2	17	134	70.28	18.35	26.12	70	27	43	123	86.31	14.56	16.87	87	20

SD standard deviation, *CV* coefficient of variation, *IQR* interquartile range

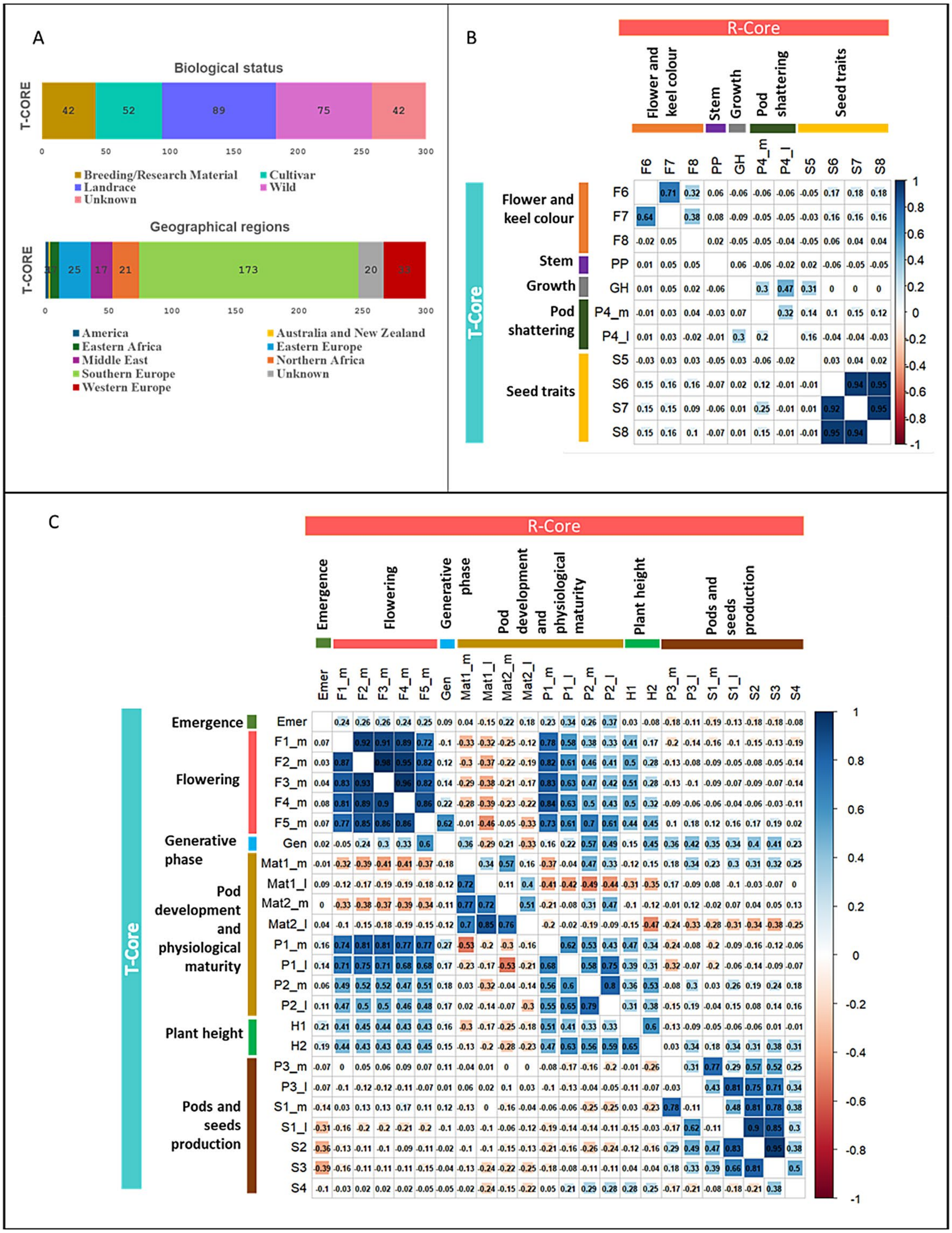

Fig. 5 Characterization and comparative phenotypic structures of the *L. albus* R-CORE and T-CORE collections. (A) Biological status and geographical region of the 300 accessions forming the T-CORE collection. (B, C) Bivariate correlation plot of (B) 11 qualitative traits and (C) 24 quantitative traits, with the upper right panel representing the R-CORE and the lower left panel representing the T-CORE (red and blue colors indicate negative and positive correlations, respectively)

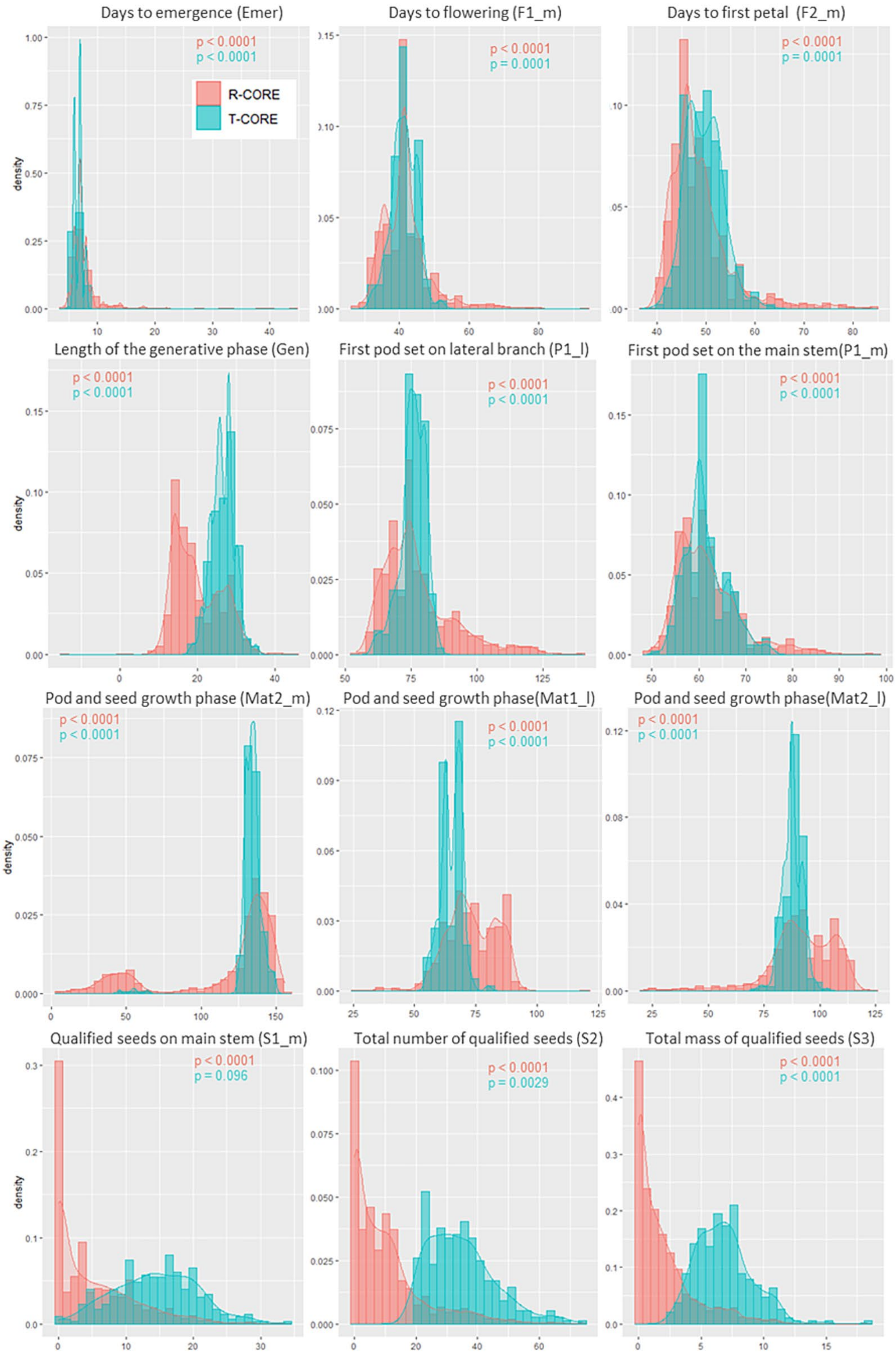

Fig. 6 Comparative quantitative variation in the R-CORE (red) and T-CORE (green) collections

Fig. 6 (continued)

PCA and HCPC were also applied while excluding both ornamentation traits and the breeding/research materials to maintain genetic integrity and to avoid skewed results caused by selection processes (Additional File 1: Table S9). PCA excluding ornamentation traits (S6–S8) clustered the accessions in three groups: I (190 accessions), II

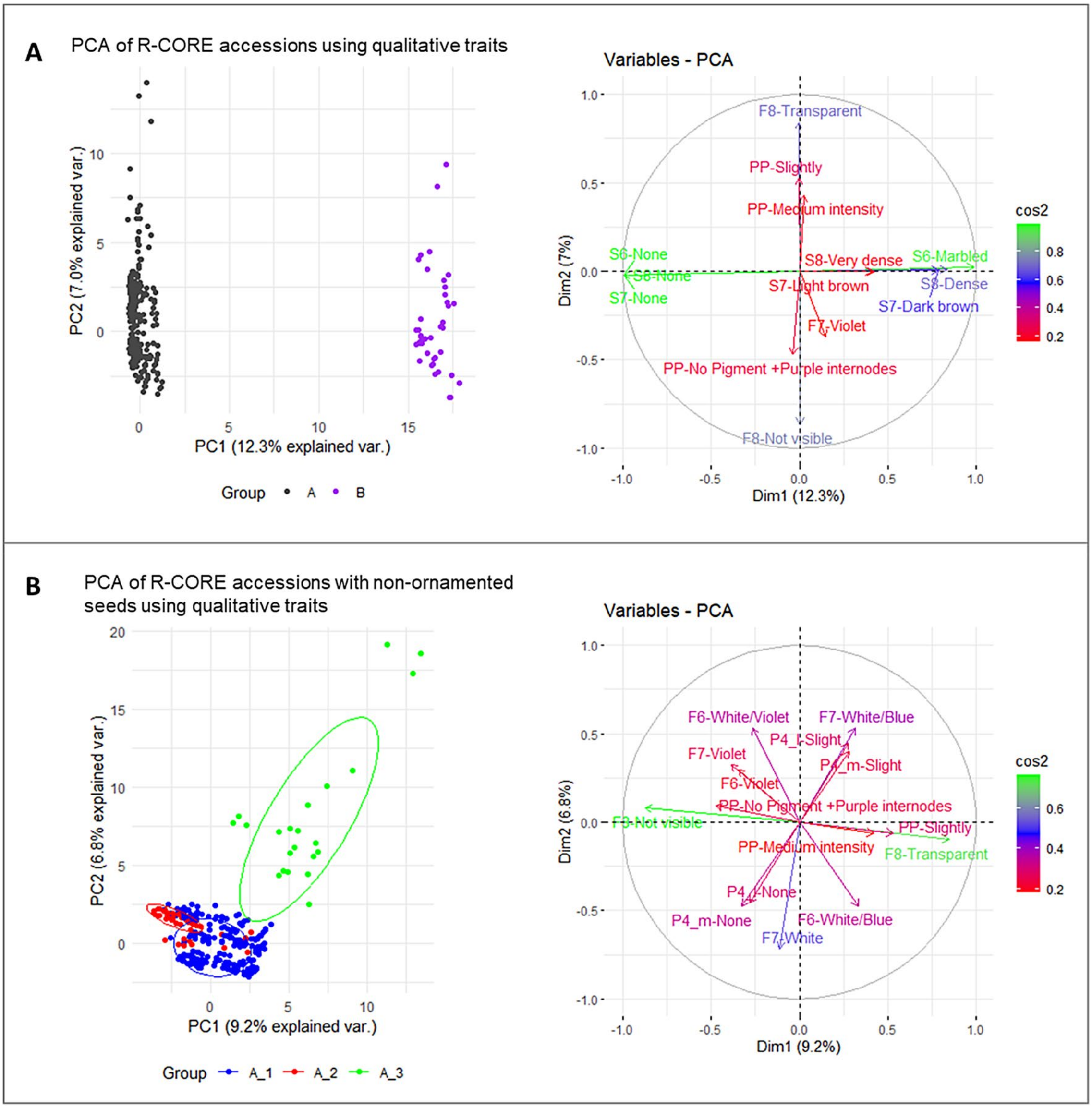

Fig. 7 Principal component analysis of the *L. albus* R-CORE accessions. (A) Biplot of all R-CORE accessions using 11 qualitative traits. (B) Biplot of R-CORE accessions excluding the accessions with ornamented seeds

(2071 accessions), and III (27 accessions) (Additional File 2: Figure S2). Cluster I consisted of accessions with violet, white/pink or pink flowers, an invisible keel, semi-determinate growth, and no pigmentation on the main stem. Cluster II consisted of accessions with white/blue, white or white/violet flowers, non-shattering pods, a transparent keel, and purple pigmentation on main stem. Cluster III was differentiated mainly by accessions with shattering pods and a white/transparent keel (Additional File 1: Table S9).

The exclusion of breeding/research materials also resulted in three clusters (I, II and III) (Additional File 1: Table S10, Additional File 2: Figure S3). Cluster I was differentiated by an invisible keel, no pigmentation on the main stem, white/violet, white or violet flowers, non-ornamented seeds, and semi-determinate growth. Cluster II consisted of accessions with a transparent/white keel, purple pigmentation on the main stem, indeterminate growth, and (uniquely) shattering pods. Cluster III included accessions with ornamented seeds, pink or white/pink flowers, and determinate growth. When the ornamentation traits were excluded, PCA again resulted in three groups (I, II and III) (Additional 2: Figure S4). Cluster I featured accessions with violet, white/pink or pink flowers, an invisible keel and semi-determinate growth. Cluster II featured accessions with white/blue, white or white/violet flowers, non-shattering pods, and a transparent keel. Cluster III was characterized by shattering pods and a white/transparent keel (Additional File 1: Table S10).

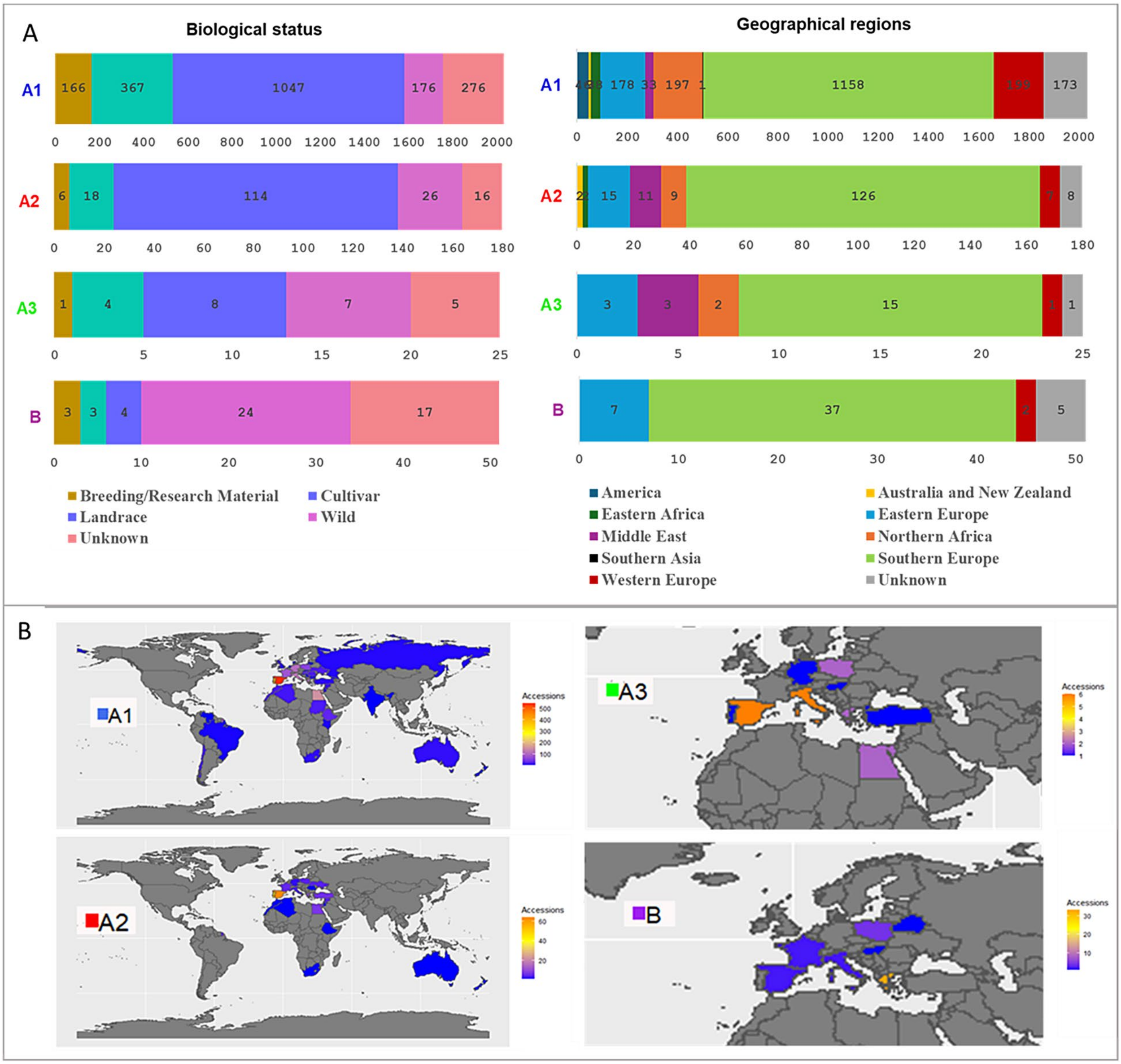

Fig. 8 Clustering of the R-CORE accessions in groups A1, A2, A3, and B. (A) Biological status and geographical distribution. (B) Country of origin

PCA was also applied to the 24 quantitative traits to identify those differentiating between geographical regions, biological status and defined phenotypic groups. The two first principal components explained 38.3% and 24.5% of the total variance, respectively (Additional File 2: Figure S5, Additional File 1: Table S11). However, the biplot showed no distinct separation of R-CORE accessions based on phenotypic groups, geographical regions or biological status (Additional File 2: Figure S5).

Association of flowering and seed traits with phenotypic groups, geographical origin and biological status

Comparative ANOVA was applied to important domestication traits such as flowering time and seed characteristics, revealing that accessions in group A were early flowering in general, and group A3 showed more significant earlier flowering (40 days) than group B (47 days). Group A3 was also significantly more productive (~25 qualified seeds) compared to groups A1, A2 and B (~11, eight and 11 seeds, respectively) (Additional File 2: Figures S6, S7). All accessions from the Middle East were early flowering and the most productive, with an average of 39 days to first bud occurrence (F6) and an average of 20 qualified seeds (Additional File 2: Figures S8 and S9). In contrast, accessions from Australia and New Zealand, America, and Southern Europe were late flowering (43 days) and the least productive, with an average of seven, nine and 10 seeds, respectively. In terms of biological status, landraces and wild accessions were later flowering

(42 days) than breeding materials and cultivars (~ 40 days) (Additional File 2: Figure S10).

The ornamented accessions (group B) were characterized by a significantly lower mean 1000 seed weight (85.01 g) than group A (145.71 g), featuring non-ornamented accessions. Among the latter, the 1000 seed weight was significantly higher in group A3 (pod-shattering accessions; 203.25 g) than groups A1 and A2 (145.16 and 136.33, respectively) (Additional File 2: Figure S7). The 1000 seed weight also differed significantly between geographical regions. The highest values were observed in accessions from Northern Africa (196.09 g) and the Middle East (171.95 g) and the lowest in those from Australia and New Zealand (86.96 g) (Additional File 2: Figure S9). There were no significant differences in 1000 seed weight between breeding/research materials (153.02 g), landraces (146.82 g) and cultivars (140.69 g) (Additional File 2: Figure S11).

We distinguished a group of 209 wild accessions with non-ornamented seeds (Wild_A) and a group of 24 with ornamented seeds (Wild_B) (Fig. 9A). Wild_A accessions originated from different regions, whereas Wild_B accessions were mainly from Greece. Notably, the Wild_B group was significantly later flowering (47 days) and produced fewer seeds (11) with a lower 1000 seed weight (85.01 g) than the Wild_A group, which flowered after 40 days and produced an average of 25 seeds with a 1000 seed weight of 185.27 g (Fig. 9B).

Patterns of total genetic variation

The CV_g was highly variable across traits, phenotypic groups, geographical regions, and biological statuses (Table 7, Additional File 1: Table S12). Traits related to the production of pods and seeds, such as the number of qualified seeds on the main stem (S1_m), total mass (S3), the total number of qualified seeds (S2), 1000 seed weight (S4), and the number of pods per plant on the main stem (P3_m), had the highest CV_g (> 30%, average 34%), whereas the flowering traits had the lowest CV_g with an average of 5%. Regarding the defined phenotypic groups, the mean CV_g for all traits was highest in the phenotypic group of non-ornamented accessions (group A). Group A3 (with pod-shattering accessions) was the highest of all (CV_g = 17%), followed by groups A1 (CV_g = 15%) and A2 (CV_g = 14%). In terms of biological status, the highest overall CV_g was observed for breeding/research materials (19%) and Wild_A (18%), followed by landraces (14%), cultivars (13%) and Wild_B (10%). At the geographical level, the overall total genetic variation was highest in Eastern Europe (CV_g = 23%), Eastern Africa (CV_g = 21%) and the Middle East (CV_g = 21%), and lowest in Northern Africa and Western Europe (CV_g = 12%) (Additional File 1: Table S12).

Discussion

Crops genetic diversity is crucial to develop sustainable, resilient agricultural systems [58]. The development, characterization and maintenance of PGRs to preserve genetic diversity will ensure their availability for crop improvement, especially to mitigate the effects of climate change and address the global demand for healthy and nutritious food. We developed a nested core collection of *L. albus* germplasm that significantly expands the geographic and genetic scope of previous collections, which

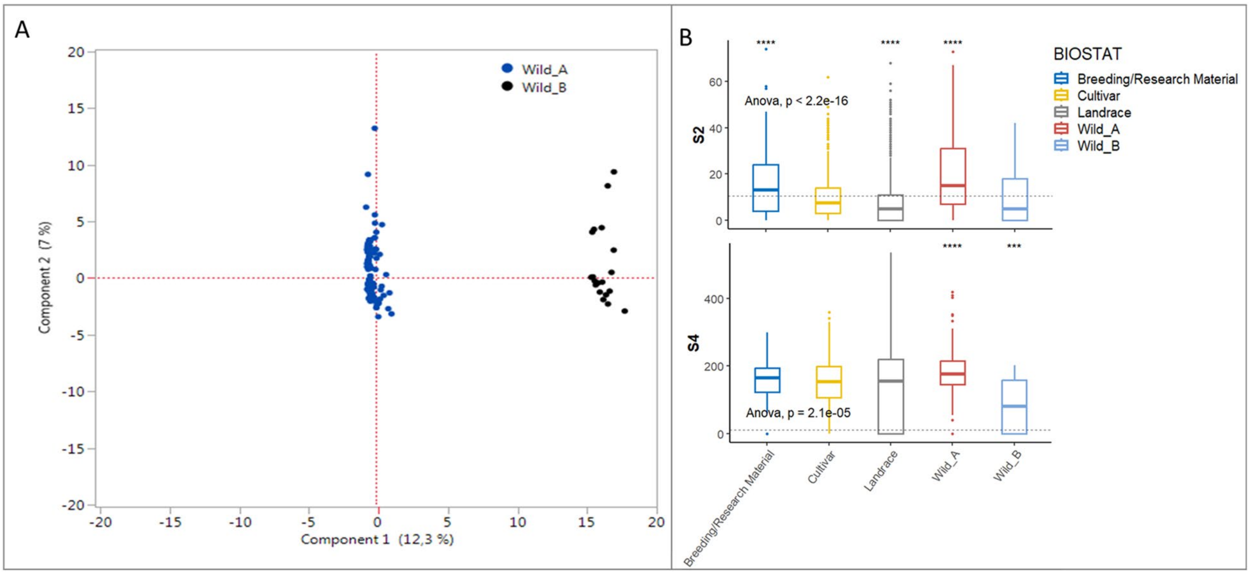

Fig. 9 Analysis of wild accessions. (A) Phenotypic structure of wild accessions. (B) Box plots showing differences in the total number of qualified seeds (S2) and 1000 seed weight (S4) compared to accessions with a different biological status. Statistical significance was determined by one-way ANOVA followed by Tukey's multiple comparisons test (*$p \leq 0.01$)

Table 7 Patterns of total genetic variation among the 24 quantitative traits in terms of phenotypic groups and the biological status of the accessions

Phenotype	**Traits**	**Phenotypic groups**				**Biological status**				
		A1	**A2**	**A3**	**B**	**Breeding/research material**	**Cultivar**	**Landrace**	**Wild_A**	**Wild_B**
		CV_g %	**CV_g %**	**CV_g %**	**CV_g %**	**CV_g %**	**CV_g %**	**CV_g %**	**CV_g %**	**CV_g %**
Emergence	*Emer*	4.00	6.00	8.00	41.00	7.00	8.00	4.00	12.00	2
Flowering traits	*F1_m*	9.00	7.00	10.00	6.00	8.00	8.00	8.00	11.00	0
	F2_m	8.00	8.00	7.00	5.00	6.00	5.00	7.00	8.00	4
	F3_m	7.00	8.00	7.00	6.00	6.00	5.00	6.00	7.00	2
	F4_m	7.00	7.00	6.00	6.00	5.00	5.00	6.00	7.00	2
	F5_m	6.00	5.00	6.00	2.00	5.00	4.00	5.00	6.00	0
	Gen	2.00	3.00	2.00	13.00	8.00	8.00	4.00	6.00	12
Average for flowering traits		6.50	6.33	6.33	6.33	6.33	5.83	6.00	7.50	3.33
Pod development and physiological maturity	*P1_m*	6.00	0.00	5.00	11.00	5.00	4.00	4.00	6.00	9
	P1_l	4.00	12.00	6.00	0.00	4.00	5.00	4.00	8.00	0
	P2_m	4.00	3.00	3.00	13.00	4.00	3.00	3.00	3.00	0
	P2_l	2.00	1.00	2.00	5.00	2.00	2.00	2.00	3.00	0
	Mat1_m	30.00	29.00	27.00	0.00	17.00	24.00	32.00	33.00	0
	Mat1_l	38.00	31.00	30.00	19.00	15.00	22.00	39.00	37.00	14
	Mat2_m	25.00	41.00	26.00	0.00	22.00	0.00	24.00	32.00	0
	Mat2_l	25.00	31.00	34.00	0.00	54.00	7.00	18.00	33.00	0
Average for pod development and physiological maturity		16.75	18.50	16.63	6.00	15.38	8.38	15.75	19.38	2.87
Pods and seed production	*P3_m*	21.00	17.00	29.00	20.00	24.00	21.00	16.00	33.00	26
	P3_l	28.00	24.00	28.00	0.00	38.00	24.00	16.00	32.00	18
	S1_m	29.00	23.00	40.00	53.00	49.00	33.00	40.00	35.00	52
	S1_l	27.00	10.00	31.00	0.00	45.00	31.00	16.00	32.00	0
	S2	23.00	5.00	26.00	33.00	37.00	25.00	21.00	25.00	35
	S3	25.00	9.00	29.00	30.00	40.00	29.00	28.00	23.00	38
	S4	16.00	4.00	15.00	0.00	16.00	18.00	13.00	13.00	0
Average for pods and seeds production traits		24.14	13.14	28.29	19.43	35.57	25.86	21.43	27.57	24.14
Plant height	*H1*	11.00	28.00	13.00	15.00	13.00	13.00	12.00	14.00	18
	H2	13.00	17.00	15.00	13.00	18.00	16.00	13.00	14.00	12
Average for plant height traits		12.00	22.50	14.00	14.00	15.50	14.50	12.50	14.00	15
Average for all traits		15.42	13.71	16.88	12.13	18.67	13.33	14.21	18.04	10.17

were largely derived from Southern Europe, mainly Portugal [59], the northwest Iberian plateau [41], Greece [39], and Ethiopia [33, 44], and typically encompassed a small number of accessions. The effective utilization of PGRs is often hindered by extensive, uncharacterized collections, whereas a smaller, well-curated subset that maximizes diversity and breeding relevance could facilitate their more efficient utilization. Therefore, T-CORE is a multipurpose resource that has fewer accessions than the R-CORE but still maximizes genetic diversity and representation [2, 4], saving time, labor and costs during subsequent analysis. The lupin T-CORE contains ~ 13% of the R-CORE accessions (2288), as similarly reported for other legumes, including lentil [19, 20], chickpea [60], and common bean [61]. The variation in our T-CORE collection was representative of the R-CORE, as confirmed by the conserved patterns of correlations between the collections for all combinations of quantitative and qualitative traits, as reported in previous studies involving crops such as lentil [18], chickpea [60], wheat [62], and eggplant [63]. We observed significant variations in many quantitative and qualitative agro-morphological traits reflecting the diverse nature of the collection comprising wild accessions, landraces, breeding materials, and advanced cultivars from a wide range of geographical origins [62]. Notably, the accessions with ornamented seeds formed a distinct and interesting group (group B) that may be helpful to understand the evolutionary events in lupins. Accessions from this group formed a subgroup (Wild_B), originating from Greece, marked by fewer seeds and a low seed weight —traits typical of wild types adapted for dispersal. This group may represent a wild ancestor ("graecus type") of *Lupinus albus*, as Greece is likely its center of origin [23]. Moreover, considerable variation in the seed ornamentation traits, particularly in secondary seed color has not been described before [41] and the loss of seed pigmentation is a sign of domestication syndrome [64, 65]. These accessions could also help to determine the impact of seed coat pigmentation on root colonization by plant-growth-promoting microbes [66], which would contribute significantly to overall soil health by promoting symbiotic relationships with *Rhizobium* spp. in diverse agroecosystems. Similarly, some wild accessions in group A3 with indeterminate growth, high seed weight and shattering pods can be suitable for studies on the origin and domestication of *L. albus*, representing untapped wild material that could be exploited to increase yields. For example, the presence of non-shattering pods is a primary trait associated with domestication syndrome in legumes [25, 64, 65, 67, 68, 69, 70], yet the indeterminate accessions could be useful for targeting yield traits even though breeders usually prefer determinate accessions, which are more suitable for mechanical harvesting. With regard to quantitative traits, the *L. albus* accessions did not cluster into defined groups or according to their passport data, suggesting a weak population structure, which is consistent with earlier reports on *L. albus* [33, 71, 72, 73] and other lupins [74, 75]. This may reflect the slow and sporadic domestication of *L. albus* [71, 73].

Nevertheless, the phenotypic data provided in this study will be useful for the breeders to identify the accessions for a preferable trait such as early emerging, early flowering and early maturing lines. For instance, our study highlights that Ethiopian *L. albus* accessions form a unique and important gene pool, with some wild and landrace types from Eastern Africa, particularly Ethiopia, showing early flowering and higher productivity. These wild and landrace accessions represent an untapped genetic reservoir and could be a valuable resource for breeding programs, supporting previous findings [33, 39, 44, 76].

The observed genetic variation may also reflect the history of *L. albus* domestication and adaptation during evolution. The low total genetic variation observed for flowering traits may indicate that many beneficial alleles for flowering traits arose during domestication [76]. This would deplete the genetic variation available during selection, as discussed for maize [77], but further research is needed to test this assumption. However, moderate to high variation in traits related to physiological maturity and seed production indicate these traits remain under substantial genetic control and may respond well to selection [78]. This highlights their relevance for adaptation in the context of climate change [79]. In general, favorable environments are associated with weak selection that increases genetic variation [80]. We observed this pattern for total genetic variation in accessions from Eastern Europe, Eastern Africa and the Middle East, indicating that these regions may be the most favorable for growing *L. albus*. Notably, landraces showed the least total genetic variation, possibly indicating that the environmental stress caused by climate change depletes the additive genetic variation in landraces, which may affect their evolutionary potential [81].

The conservation of SSD collections reduces the likelihood of genetic changes during seed multiplication for heterogeneous accessions in gene banks (i.e., genetic drift and/or selection), and standardized phenotyping protocols used in this study reduce data bias, enhancing genotype-phenotype association [1, 25]. Our model approach for the development of nested core collections allows the construction of *Intelligent Collections* based on SSD lines, which enables the effective use of lupin genetic resources [25], in parallel with extensively studied legumes such as common bean, chickpea, lentil [46, 82, 83, 84, 85, 86]. The presented core collections are a unique source for the exploitation of phenomics for gene bank material exploration but can be also used by specific users to develop and exploit germplasm at different levels of genotypic and phenotypic characterization. To facilitate the proper exploration, management and sharing of genetic resources, we have assigned the DOIs for some accessions (1178), for the first time for lupin genetic resources at this scale, and as mentioned above this approach is unique in comparisons with other already published core collections. Of note, once the project is finished, the distribution of genetic resources is anticipated to follow the approach described by Bellucci et al. [1], and the data will be searchable and filterable in the INCREASE web portal (https://www.pulsesincrease.eu/). Furthermore, the T-CORE will be tested also in the field environment providing both standard and molecular phenotyping through metabolic and gene expression data. Integrating environmental data with field performance observations and other -omic data will then be crucial to refine genomic-enabled prediction models and improve their accuracy. Integrating present studied along with genomic variation with environment-of-origin data through genome–environment associations (GEAs) and environmental genomic selection (EGS), will help to identify the resilient and adaptable lupin genotypes for future climates as it was shown for sorghum [87, 88]. Integrating ecophysiological modeling, remote sensing, and machine learning can further enhance EGS, promoting sustainability through optimized resource use [87]. Overall, the *L. albus* nested core collections developed in this study can help to advance research on the evolution [82, 89], genomic selection [90, 91] and genetic architecture of important agronomic traits [92, 93] such as grain yield [94].

Breeding of crops that can tolerate extreme climates is vital, a process that requires characterizing, utilizing and mobilizing diverse genetic resources. Identifying the phenotypic variation and construction of the core collections provide a more comprehensive strategy of diversifying breeding for global food security [87]. In this study, we focus on lupins, to promote the lupin improvement initiatives, especially regarding the limited access to lupin germplasm, and its scared characterization. Moreover, the approach presented here might be useful for the development of genetic resources and crop improvement programs for orphan legume crops [95].

Conclusions

This study is the first step towards the development and characterization of *Intelligent Collections* of *L. albus* genetic resources to conserve and manage agrobiodiversity. The R-CORE collection is the largest, best-characterized, and most diverse genetic resource for *L. albus*, supporting agricultural resilience, productivity, and studies on evolution and domestication in this species. We also developed a smaller, multipurpose T-CORE collection designed to streamline and accelerate the use of *Lupinus* genetic resources in crop improvement programs. These collections provide a strong foundation for studying key agronomic traits such as early flowering, seed production, grain yield, flower color, and seed morphology, and will be useful for a wide range of applications from basic research to targeted breeding and genomic selection.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12870-025-07115-3.

Supplementary Material 1

Acknowledgements

This study was conducted as part of the INCREASE project. INCREASE has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no. 862862. This publication reflects only the author's view and neither the Research Executive Agency (REA) nor the European Commission are responsible for any use that may be made of the information it contains.

Author contributions

UKT analyzed the data and drafted the manuscript. MT carried out phenotyping at IPG-PAN. MT, KC, HJ, MK and MN carried out greenhouse experiments at IPG-PAN, Poland, and collected the data. MN transformed the raw data in to standard format and analyzed the initial dataset. BP and VM carried out greenhouse experiments at KIS, Slovenia, and collected the data. LR, AP, EB and EB provided critical suggestions at various stages of data analysis and interpretation. MO, a genebank data expert, contributed for the data management and integration. KS and MK collected the research material from various gene banks/providers. KS and RP conceived and designed the study. MK, RP and KS supervised the research, interpreted the results, and contributed to manuscript writing and editing. RP coordinated the overall study as a part of the INCREASE project. All authors have read and approved of the final manuscript.

Funding

This study was conducted as part of the INCREASE project (Intelligent Collections of Food Legumes Genetic Resources for European Agrofood Systems) funded by the European Union Horizon 2020 Research and Innovation Programme under grant agreement no. 862862.

Data availability

Seed material is made available under the agreement governing the Stakeholder Consortium of the INCREASE project. However, once the project is completed, the core collections will be preserved as units in one or more gene banks that participate in the project, for long-term and active conservation [1]. The associated data will be searchable and filterable in the INCREASE web portal (https://www.pulsesincrease.eu/).

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 22 April 2025 / Accepted: 23 July 2025

References

1. Bellucci E, Mario Aguilar O, Alseekh S, Bett K, Brezeanu C, Cook D, De la Rosa L, Delledonne M, Dostatny DF, Ferreira JJ, et al. The INCREASE project: intelligent collections of food-legume genetic resources for European agrofood systems. Plant J. 2021;108(3):646–60.
2. McCouch S, Navabi ZK, Abberton M, Anglin NL, Barbieri RL, Baum M, Bett K, Booker H, Brown GL, Bryan GJ, et al. Mobilizing crop biodiversity. Mol Plant. 2020;13(10):1341–4.
3. Brown A. Core collections: a practical approach to genetic resources management. Genome. 1989;31(2):818–24.
4. Frankel OH. Plant genetic resources today: a critical appraisal. Crop Genetic Resources: Conserv Evaluation 1984:241–56.
5. Van Hintum TJ, Brown A, Spillane C. Core collections of plant genetic resources. Bioversity International; 2000.
6. Bellucci E, Benazzo A, Xu C, Bitocchi E, Rodriguez M, Alseekh S, Di Vittori V, Gioia T, Neumann K, Cortinovis G, et al. Selection and adaptive introgression guided the complex evolutionary history of the European common bean. Nat Commun. 2023;14(1):1908.
7. Lauterberg M, Tschiersch H, Papa R, Bitocchi E, Neumann K. Engaging precision phenotyping to scrutinize vegetative drought tolerance and recovery in C chickpea plant genetic resources. Plants. 2023;12(15):2866.
8. Jurado M, García-Fernández C, Campa A, Ferreira JJ. Genetic erosion within the Fabada dry bean market class revealed by high-throughput genotyping. Plant Genome. 2023;16(4):e20379.
9. Odong T, Jansen J, Van Eeuwijk F, van Hintum TJ. Quality of core collections for effective utilisation of genetic resources review, discussion and interpretation. Theor Appl Genet. 2013;126:289–305.
10. Oliveira MF, Nelson RL, Geraldi IO, Cruz CD, de Toledo JFF. Establishing a soybean germplasm core collection. Field Crops Res. 2010;119(2–3):277–89.

11. Blair MW, Díaz LM, Buendía HF, Duque MC. Genetic diversity, seed size associations and population structure of a core collection of common beans (*Phaseolus vulgaris* L.). Theor Appl Genet. 2009;119:955–72.
12. García-Fernández C, Jurado M, Campa A, Brezeanu C, Geffroy V, Bitocchi E, Papa R, Ferreira JJ. A core set of snap bean genotypes established by phenotyping a large panel collected in Europe. Plants. 2022;11(5):577.
13. Mahalakshmi V, Ng Q, Lawson M, Ortiz R. Cowpea [*Vigna unguiculata* (L.) Walp.] core collection defined by geographical, agronomical and botanical descriptors. Plant Genetic Resour. 2007;5(3):113–9.
14. Reddy L, Upadhyaya HD, Gowda C, Singh S. Development of core collection in Pigeonpea [*Cajanus cajan* (L.) Millspaugh] using geographic and qualitative morphological descriptors. Genet Resour Crop Evol. 2005;52:1049–56.
15. Upadhyaya H, Reddy K, Gowda C, Singh S. Phenotypic diversity in the Pigeonpea (*Cajanus cajan*) core collection. Genet Resour Crop Evol. 2007;54:1167–84.
16. Upadhyaya HD, Ortiz R, Bramel PJ, Singh S. Development of a groundnut core collection using taxonomical, geographical and morphological descriptors. Genet Resour Crop Evol. 2003;50:139–48.
17. Upadhyaya HD, Bramel PJ, Singh S. Development of a Chickpea core subset using geographic distribution and quantitative traits. Crop Sci. 2001;41(1):206–10.
18. Tripathi K, Kumari J, Gore PG, Mishra DC, Singh AK, Mishra GP, Gayacharan C, Dikshit H, Singh N, Semwal D, et al. Agro-morphological characterization of lentil germplasm of Indian National genebank and development of a core set for efficient utilization in lentil improvement programs. Front Plant Sci. 2022;12:751429.
19. Tullu A, Kusmenoglu I, McPhee K, Muehlbauer F. Characterization of core collection of lentil germplasm for phenology, morphology, seed and straw yields. Genet Resour Crop Evol. 2001;48:143–52.
20. Simon CJ, Hannan RM. Development and use of core subsets of cool-season food legume germplasm collections. HortScience. 1995;30(4):C907–907.
21. Vaijayanthi P, Ramesh S, Gowda MB, Rao AM, Keerthi C. Development of core sets of dolichos bean (*Lablab purpureus* L. Sweet) germplasm. J Crop Improv. 2015;29(4):405–19.
22. Vilakazi B, Mafongoya PL, Odindo AO, Phophi MM. The role of neglected grain legumes in food and nutrition security and human health. Sustainability. 2025;17(1):350.
23. Gladstones JS. Distribution, origin, taxonomy, history and importance. In: *Lupins as crop plants: biology, production, and utilization.* Edited by Gladstones JS, Atkins CA, Hamblin J: CAB International; 1998: 1–36.
24. Nartea A, Kuhalskaya A, Fanesi B, Orhotohwo OL, Susek K, Rocchetti L, Di Vittori V, Bitocchi E, Pacetti D, Papa R. Legume byproducts as ingredients for food applications: preparation, nutrition, bioactivity, and techno-functional properties. Compr Rev Food Sci Food Saf. 2023;22(3):1953–85.
25. Kroc M, Tomaszewska M, Czepiel K, Bitocchi E, Oppermann M, Neumann K, Guasch L, Bellucci E, Alseekh S, Graner A, et al. Towards development, maintenance, and standardized phenotypic characterization of single-seed-descent genetic resources for lupins. Curr Protocols. 2021;1(7):e191.
26. de Visser C, Schreuder R, Stoddard F. The EU's dependence on soya bean import for the animal feed industry and potential for EU produced alternatives. *Oilseeds and fats, crops and lipids* 2014.
27. Yeheyis L, Sergeant A, Nelson M, Mcnaughton D, Sanders H. Opportunities and challenges for market oriented lupin (*Lupinus* spp.) production in Ethiopia. Turkish J Agriculture-Food Sci Technol. 2022;10(6):1165–73.
28. Susek K, Vincenzi L, Tomaszewska M, Kroc M, Franco E, Cosentino E, Limongi AR, Tanwar UK, Jamil H, Nelson MN, et al. The unexplored diversity of rough-seeded lupins provides rich genomic resources and insights into lupin evolution. Nat Commun. 2025;16(1):1–16.
29. Martinez-Hernandez JE, Salvo-Garrido H, Levicoy D, Caligari PD, Rupayán A, Moyano T, Carrasco M, Hernandez S, Armijo-Godoy G, Westermeyer F, et al. Genomic structure of yellow lupin (*Lupinus luteus*): genome organization, evolution, gene family expansion, metabolites and protein synthesis. BMC Genomics. 2025;26(1):1–20.
30. Pancaldi F, Gulisano A, Severing EI, van Kaauwen M, Finkers R, Kodde L, Trindade LM. The genome of *Lupinus mutabilis*: evolution and genetics of an emerging bio-based crop. Plant J. 2024;120(3):881–900.
31. Tosoroni A, Di Vittori V, Nanni L, Musari E, Papalini S, Bitocchi E, Bellucci E, Pieri A, Ghitarrini S, Susek K, et al. Recent advances in molecular tools and pre-breeding activities in white lupin (*Lupinus albus*). Plants. 2025;14(6):914.
32. Musco N, Cutrignelli MI, Calabrò S, Tudisco R, Infascelli F, Grazioli R, Lo Presti V, Gresta F, Chiofalo B. Comparison of nutritional and antinutritional traits among different species (*Lupinus albus* L., *Lupinus luteus* L., *Lupinus angustifolius* L.) and varieties of lupin seeds. J Anim Physiol Anim Nutr. 2017;101(6):1227–41.
33. Atnaf M, Yao N, Martina K, Dagne K, Wegary D, Tesfaye K. Molecular genetic diversity and population structure of Ethiopian white lupin landraces: implications for breeding and conservation. PLoS ONE. 2017;12(11):e0188696.
34. Gresta F, Wink M, Prins U, Abberton M, Capraro J, Scarafoni A, Hill G. Lupins in European cropping systems. In: *Legumes in cropping systems*. CABI Wallingford UK; 2017: 88–108.
35. Ehab E-H, Ashrie A, Ammar M, Alghamdi S. Genetic variation among Egyptian white lupin (*Lupinus albus* L.) genotypes. Turkish J Field Crops. 2016;21(1):148–55.
36. Hibstu A. Genetic diversity and association of traits in white lupin (*Lupinus albus* L.) accessions of Ethiopia. Doctoral dissertation, Haramaya University; 2016.
37. Mulugeta A, Kassahun T, Dagne K, Dagne W. Extent and pattern of genetic diversity in Ethiopian white lupin landraces for agronomical and phenological traits. Afr Crop Sci J. 2015;23(4):327–41.
38. Berger JD, Shrestha D, Ludwig C. Reproductive strategies in mediterranean legumes: trade-offs between phenology, seed size and vigor within and between wild and domesticated *Lupinus* species collected along aridity gradients. Front Plant Sci. 2017;8:548.
39. Zafeiriou I, Polidoros AN, Baira E, Kasiotis KM, Machera K, Mylona PV. Mediterranean white lupin landraces as a valuable genetic reserve for breeding. Plants. 2021;10(11):2403.
40. Lagunes-Espinoza L, Huyghe C, Papineau J. Genetic variation for pod wall proportion in *Lupinus albus*. Plant Breeding. 2000;119(5):421–5.
41. Gonzalez-Andres F, Casquero PA, San-Pedro C, Hernández-Sánchez E. Diversity in white lupin (*Lupinus albus* L.) landraces from Northwest Iberianplateau. Genet Resour Crop Evol. 2007;54:27–44.
42. Tizazu H, Emire SA. Chemical composition, physicochemical and functional properties of lupin (*Lupinus albus*) seeds grown in Ethiopia. Afr J Food Agric Nutr Dev 2010, 10(8).
43. Zelalem K, Chandravanshi BS. Levels of essential and non-essential elements in raw and processed *Lupinus albus* L.(White lupin, Gibto) cultivated in Ethiopia. Afr J Food Agric Nutr Dev. 2014;14(5):2015–35.
44. Beyene C. Genetic variation among white lupin (*Lupinus albus* L.) landraces from Northwestern and Southern Ethiopia for agronomic traits and nutrient contents of grain. J Plant Breed Crop Sci. 2020;12(2):156–69.
45. Parra-Quijano M, Iriondo JM, de la Cruz M, Torres E. Strategies for the development of core collections based on ecogeographical data. Crop Sci. 2011;51(2):656–66.
46. Cortinovis G, Oppermann M, Neumann K, Graner A, Gioia T, Marsella M, Alseekh S, Fernie AR, Papa R, Bellucci E, et al. Towards the development, maintenance, and standardized phenotypic characterization of single-seed-descent genetic resources for common bean. Curr Protocols. 2021;1(5):e133.
47. Rocchetti L, Bellucci E, Cortinovis G, Di Vittori V, Lanzavecchia G, Frascarelli G, Nanni L, Del Gatto A, Pieri S, Mangoni L, et al. The development of a European and mediterranean chickpea association panel (EMCAP). Agronomy. 2020;10(9):1417.
48. Lê S, Josse J, Husson F. FactoMineR: an R package for multivariate analysis. J Stat Softw. 2008;25:1–18.
49. Yilan L, Rutong Z, Clustertend. Check the Clustering Tendency, R Package Version 1.4. 2015.
50. Alboukadel K, Mundt F, Factoextra. Extract and visualize the results of multivariate data analyses. R Package Version 1.0.5. 2017.
51. Bates D, Maechler M, Bolker B, Walker S, Christensen RHB, Singmann H, Dai B, Grothendieck G, Green P,Bolker MB. Package 'lme4'.convergence2015, 12(1):2.
52. Garcia-Gonzalez F, Simmons LW, Tomkins JL, Kotiaho JS, Evans JP. Comparing evolvabilities: common errors surrounding the calculation and use of coefficients of additive genetic variation. Evolution. 2012;66(8):2341–9.
53. Aravind J, Kaur V, Wankhede D, Nanjundan J. EvaluateCore: quality evaluation of core collections. R Package Version. 2022;3:01.
54. Kim K-W, Chung H-K, Cho G-T, Ma K-H, Chandrabalan D, Gwag J-G, Kim T-S, Cho E-G, Park Y-J. PowerCore: a program applying the advanced M strategy with a heuristic search for Establishing core sets. Bioinformatics. 2007;23(16):2155–62.
55. Shannon CE, Weaver W. The mathematical theory of communication. University of Illinois Press; 1949.
56. Wei T, Simko V, Levy M, Xie Y, Jin Y, Zemla J. Package 'corrplot'. Statistician. 2017;56(316):e24.
57. Meyer D, Zeileis A, Hornik K. vcd: visualizing categorical data. R package version 1.4–8. 2020.

58. Santamarina C, Mathieu L, Bitocchi E, Pieri A, Bellucci E, Di Vittori V, Susek K, Scossa F, Nanni L, Papa R. Agroecological genomics and participatory science: optimizing crop mixtures for agricultural diversification. Trends Plant Sci. 2025. https://doi.org/10.1016/j.tplants.2025.04.008
59. Vaz AC, Pinheiro C, Martins JM, Ricardo CP. Cultivar discrimination of Portuguese *Lupinus albus* by seed protein electrophoresis: the importance of considering glutelins and glycoproteins. Field Crops Res. 2004;87(1):23–34.
60. Archak S, Tyagi RK, Harer P, Mahase L, Singh N, Dahiya OP, Nizar MA, Singh M, Tilekar V, Kumar V, et al. Characterization of chickpea germplasm conserved in the Indian National genebank and development of a core set using qualitative and quantitative trait data. Crop J. 2016;4(5):417–24.
61. Rivera A, Plans M, Sabaté J, Casañas F, Casals J, Rull A, Simó J. The Spanish core collection of common beans (*Phaseolus vulgaris* L.): an important source of variability for breeding chemical composition. Front Plant Sci. 2018;9:1642.
62. Phogat B, Kumar S, Kumari J, Kumar N, Pandey AC, Singh T, Kumar S, Tyagi R, Jacob SR, Singh AK, et al. Characterization of wheat germplasm conserved in the Indian National genebank and establishment of a composite core collection. Crop Sci. 2021;61(1):604–20.
63. Gangopadhyay K, Mahajan R, Kumar G, Yadav S, Meena B, Pandey C, Bisht I, Mishra S, Sivaraj N, Gambhir R, et al. Development of a core set in Brinjal (*Solanum melongena* L). Crop Sci. 2010;50(3):755–62.
64. Bohra A, Kilian B, Sivasankar S, Caccamo M, Mba C, McCouch SR, Varshney RK. Reap the crop wild relatives for breeding future crops. Trends Biotechnol. 2022;40(4):412–31.
65. Smýkal P, Parker T. Domestication-related changes in seed dispersal and pigmentation: visual selection and functional trait? Mol Plant. 2023;16(8):1240–2.
66. Jaiswal SK, Dakora FD. Seed-coat pigmentation plays a crucial role in partner selection and N_2 fixation in legume-root–microbe associations in African soils. Plants. 2024;13(11):1464.
67. Maity A, Lamichaney A, Joshi DC, Bajwa A, Subramanian N, Walsh M, Bagavathiannan M. Seed shattering: a trait of evolutionary importance in plants. Front Plant Sci. 2021;12:657773.
68. Rau D, Murgia ML, Rodriguez M, Bitocchi E, Bellucci E, Fois D, Albani D, Nanni L, Gioia T, Santo D, Marcolungo L, Delledonne M, Attene G, Papa R. Genomic dissection of pod shattering in common bean: mutations at non-orthologous loci at the basis of convergent phenotypic evolution under domestication of leguminous species. Plant J. 2019;97(4):693–714.
69. Murgia ML, Attene G, Rodriguez M, Bitocchi E, Bellucci E, Fois D, Nanni L, Gioia T, Albani DM, Papa R, et al. A comprehensive phenotypic investigation of the pod-shattering syndrome in common bean. Front Plant Sci. 2017;8:251.
70. Di Vittori V, Gioia T, Rodriguez M, Bellucci E, Bitocchi E, Nanni L, Attene G, Rau D, Papa R. Convergent evolution of the seed shattering trait. Genes (Basel). 2019;10:68.
71. Hufnagel B, Soriano A, Taylor J, Divol F, Kroc M, Sanders H, Yeheyis L, Nelson M, Péret B. Pangenome of white lupin provides insights into the diversity of the species. Plant Biotechnol J. 2021;19(12):2532–43.
72. Iqbal MJ, Mamidi S, Ahsan R, Kianian SF, Coyne CJ, Hamama AA, Narina SS, Bhardwaj HL. Population structure and linkage disequilibrium in *Lupinus albus* L. germplasm and its implication for association mapping. Theor Appl Genet. 2012;125:517–30.
73. Alkemade JA, Nazzicari N, Messmer MM, Annicchiarico P, Ferrari B, Voegele RT, Finckh MR, Arncken C, Hohmann P. Genome-wide association study reveals white lupin candidate gene involved in anthracnose resistance. Theor Appl Genet. 2022;135(3):1011–24.
74. Mousavi-Derazmahalleh M, Bayer PE, Nevado B, Hurgobin B, Filatov D, Kilian A, Kamphuis LG, Singh KB, Berger JD, Hane JK, et al. Exploring the genetic and adaptive diversity of a pan-Mediterranean crop wild relative: narrow-leafed lupin. Theor Appl Genet. 2018;131:887–901.
75. Turner KG, Huang DI, Cronk QC, Rieseberg LH. Homogenization of populations in the wildflower, Texas Bluebonnet (*Lupinus texensis*). J Hered. 2018;109(2):152–61.
76. Adhikari K, Buirchell B, Yan G, Sweetingham M. Two complementary dominant genes control flowering time in albus lupin (*Lupinus albus* L). Plant Breeding. 2011;130(4):496–9.
77. Yang CJ, Samayoa LF, Bradbury PJ, Olukolu BA, Xue W, York AM, Tuholski MR, Wang W, Daskalska LL, Neumeyer MA, et al. The genetic architecture of Teosinte catalyzed and constrained maize domestication. Proc Natl Acad Sci. 2019;116(12):5643–52.
78. Wisser RJ, Fang Z, Holland JB, Teixeira JE, Dougherty J, Weldekidan T, De Leon N, Flint-Garcia S, Lauter N, Murray SC, et al. The genomic basis for short-term evolution of environmental adaptation in maize. Genetics. 2019;213(4):1479–94.
79. Hernández-Máximo E, Vargas-Hernández JJ, López-Upton J, Sánchez-Monsalvo V. Structure of genetic variation in vegetative phenology of *Cedrela odorata* L.: implications for tree breeding. New Forest 2021:1–23.
80. Etterson JR, Shaw RG. Constraint to adaptive evolution in response to global warming. Science. 2001;294(5540):151–4.
81. Mercer KL, Perales HR. Evolutionary response of landraces to climate change in centers of crop diversity. Evol Appl. 2010;3(5–6):480–93.
82. Cortinovis G, Vincenzi L, Anderson R, Marturano G, Marsh JI, Bayer PE, Rocchetti L, Frascarelli G, Lanzavecchia G, Pieri A, et al. Adaptive gene loss in the common bean pan-genome during range expansion and domestication. Nat Commun. 2024;15(1):6698.
83. Rocchetti L, Rodriguez M, Pieri A, Papalini S, De Antoni L, Vitali E, Vincze A, Brezeanu C, Bellucci E, Nanni L, et al. Landscape genomics highlights the adaptive evolution of chickpea. *bioRxiv* 2024:2024.2006. 2006.597750.
84. Guerra-García A, Gioia T, von Wettberg E, Logozzo G, Papa R, Bitocchi E, Bett KE. Intelligent characterization of lentil genetic resources: evolutionary history, genetic diversity of germplasm, and the need for well-represented collections. Curr Protocols. 2021;1(5):e134.
85. Rossato M, Marcolungo L, De Antoni L, Lopatriello G, Bellucci E, Cortinovis G, Frascarelli G, Nanni L, Bitocchi E, Di Vittori V, et al. CRISPR-Cas9-based repeat depletion for high-throughput genotyping of complex plant genomes. Genome Res. 2023;33(5):787–97.
86. Rocchetti L, Gioia T, Logozzo G, Brezeanu C, Pereira LG, la Rosa LD, Marzario S, Pieri A, Fernie AR, Alseekh S, et al. Towards the development, maintenance and standardized phenotypic characterization of single-seed-descent genetic resources for chickpea. Curr Protocols. 2022;2(2):e371.
87. Cortés AJ. Unlocking genebanks for climate adaptation. Nat Clim Change 2025:1–3.
88. Campbell Q, Castañeda-Álvarez N, Domingo R, Bishop-von Wettberg E, Runck B, Nandkangré H, Halpin-McCormick A, Fumia N, Neyhart J, Kilian B, et al. Prioritizing parents from global genebanks to breed climate-resilient crops. Nat Clim Change 2025:1–9.
89. Garcia T, Duitama J, Zullo SS, Gil J, Ariani A, Dohle S, Palkovic A, Skeen P, Bermudez-Santana CI, Debouck DG, et al. Comprehensive genomic resources related to domestication and crop improvement traits in Lima bean. Nat Commun. 2021;12(1):702.
90. Norman A, Taylor J, Edwards J, Kuchel H. Optimising genomic selection in wheat: effect of marker density, population size and population structure on prediction accuracy. G3: Genes Genomes Genet. 2018;8(9):2889–99.
91. Hu J, Chen B, Zhao J, Zhang F, Xie T, Xu K, Gao G, Yan G, Li H, Li L, et al. Genomic selection and genetic architecture of agronomic traits during modern rapeseed breeding. Nat Genet. 2022;54(5):694–704.
92. Roorkiwal M, Jarquin D, Singh MK, Gaur PM, Bharadwaj C, Rathore A, Howard R, Srinivasan S, Jain A, Garg V, et al. Genomic-enabled prediction models using multi-environment trials to estimate the effect of genotypex environment interaction on prediction accuracy in chickpea. Sci Rep. 2018;8(1):11701.
93. Weckwerth W, Ghatak A, Bellaire A, Chaturvedi P, Varshney RK. PANOMICS Meets germplasm. Plant Biotechnol J. 2020;18(7):1507–25.
94. Pang Y, Liu C, Wang D, Amand PS, Bernardo A, Li W, He F, Li L, Wang L, Yuan X, et al. High-resolution genome-wide association study identifies genomic regions and candidate genes for important agronomic traits in wheat. Mol Plant. 2020;13(9):1311–27.
95. Efficient genetic improvement. Of orphan crops cannot follow the old path. Nat Commun. 2024;15:321.

Publisher's note